DR. MOUSA ALNABHAN

أ.د. موسى النبهان

أساسيات القياس في العلوم السلوكية

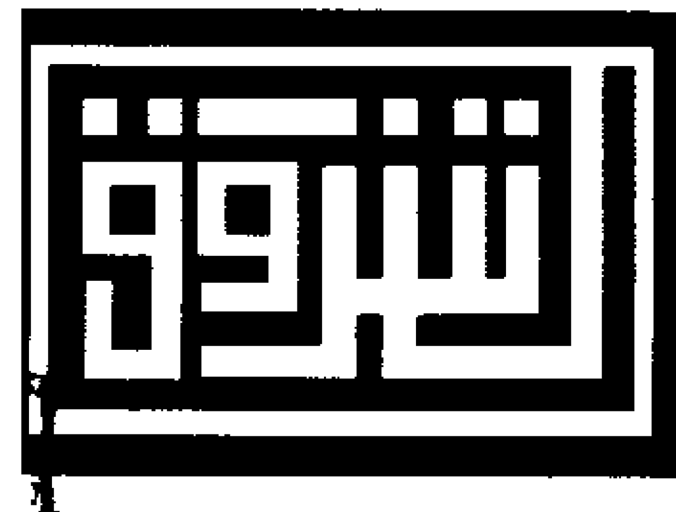


www.shorok.com

أساسيات القياس
في
العلوم السلوكية

أساسيات القياس
في
العلوم السلوكية

تأليف
الأستاذ الدكتور موسى النبهان
كلية الدراسات العليا- جامعة الخليج العربي



2013

رقم الإيداع لدى دائرة المكتبة الوطنية
(2013/1/317)

150.287

النبهان، موسى محمد

أساسيات القياس في العلوم السلوكية/ موسى محمد النبهان .ـ ط2.- عمان دار الشروق للنشر والتوزيع ,2013

(   ) ص

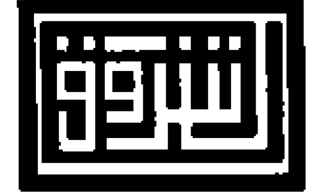
ر.إ. . 2013/1/317

الواصفات علم النفس//العلوم السلوكية//الاختبارات/

يتحمل المؤلف كامل المسؤولية القانونية عن محتوى مصنفة ولا يعبّر هذا المصنف عن رأي دائرة المكتبة الوطنية أو أي جهة حكومية أخرى.

ISBN 978-9957 - 00 - 530-6

- أساسيات القياس في العلوم السلوكية .
- تأليف : الأستاذ الدكتور موسى النبهان .
- الطبعة العربية الثانية : الإصدار الأول 2013 .
- الاحراج الداخلي وتصميم الغلاف : دائرة الإنتاج / دار الشروق للنشر والتوزيع.



دار الشروق للنشر والتوزيع

هاتف . 4618190 / 4618191 / 4624321 فاكس : 4610065

ص.ب 926463 الرمز البريدي : 11118 عمان – الاردن

Email : shorokjo@nol.com.jo

دار الشروق للنشر والتوزيع

رام الله – المصيون . نهاية شارع مستشفى رام الله

هاتف 2975632 - 2991614 - 2975633 فاكس 02/2965319

Email : shorokpr@palnet.com

الفهرس

أهدي هذا الجهد

إلى زوجتي وأبنائي عدي ومحمد ورفيف وعبدالله وشهد. فقد كانت معاناتهم تتزايد بسبب انشغالي وابتعادي عنهم غير قليل من الوقت. كما يسعدني دائما أن أتذكر بكل احترام وتقدير أستاذي الكبير: **مايكل هارول Michael Harwell** الذي لم تنقطع علاقتي به مذ تخرجت. فقد كان ومازال يعلمني، وقد وفر كثيرا من الفرص من أجلي.

لهم جميعا عميق حبي وصادق تحياتي ....

المؤلف

المؤلف في سطور

بعد حصوله على درجة البكالوريوس في الفيزياء، تخرج في جامعة بيتسبيرغ-بنسلفانيا-الولايات المتحدة الأمريكية وحصل على الماجستير عام 1988 ومن ثم الدكتوراه في منهجية البحوث والقياس والتقييم **Research Methodology, Statistics, Measurement and Assessment** عام 1991. وعمل في التدريس الجامعي في جامعات: بيتسبيرغ/الولايات المتحدة الأمريكية، ومؤتة/الأردن، والإمارات العربية المتحدة، وجامعة الخليج العربي/البحرين. كما نفذ عددا من الإستشارات العلمية والتدريبية في مجال تقييم الأداء، والذكاء العاطفي، وبناء وتقنين عدد من الإختبارات والمقاييس في أكثر من مركز في دول عربية وخليجية.

يعمل المؤلف حاليا نائبا لعميد كلية الدراسات العليا في جامعة الخليج العربي منذ 2010، وأستاذا لمقررات منهجية البحث والقياس والإحصاء في برامج الدكتوراه والماجستير في تربية الموهوبين وصعوبات التعلم والإعاقة الذهنية والتوحد منذ 2005.

عمل المؤلف أستاذا جامعيا منذ 1991 وعميدا لكلية العلوم التربوية في جامعة مؤتة (1998-2000) ومديرا للمشروع الأردني الأوروبي لبرامج إعداد المعلمين في الجامعات الأردنية الرسمية (1996-1998). كما تمتع بسنوات التفرغ العلمي التي قضاها في ربوع جامعة بيتسبيرغ /بنسلفانيا/الولايات المتحدة الأمريكية أستاذا وباحثا. وقد نشر عديدا من البحوث المتخصصة في مجلات عملية عالمية وإقليمية متخصصة باللغتين العربية والإنجليزية. كما ألف كتبا متخصصة مثل:

- **أساسيات الإحصاء في العلوم الانسانية والإجتماعية باستخدام برنامج SPSS.**
- **أساسيات دلالات فحص الفروق الكمية (الإحصاء الاستدلالي) باستخدام برنامج SPSS.**

- **منهجية البحوث الكمية والنوعية.**
- **أساليب الإحصاء المتقدم باستخدام برنامج SPSS.**
- **دليل مرجعي في الكشف عن الموهوبين.**

أشرف على أكثر من ماية رسالة ماجستير وأطروحة دكتوراه في أكثر من جامعة، كما شارك في الإشراف والمناقشات لعدد كبير من المشاريع والدراسات والورش التدريبية. إضافة إلى أنه شارك في أكثر من خمسين مؤتمر علمي إقليمي ودولي باحثا ومتحدثا ومتعلما. وقد رشح لسلسلة who is who في طبعة 2008.

مقدمة الكتاب

يحتل موضوع القياس أهمية مميزة في دراسة الظواهر والتنبؤ بتغيراتها. ويكون على أهمية أكبر عندما يتعلق بالعلوم السلوكية عامة، وبالسلوك الإنساني تحديدا. ذلك بسبب طبيعة ذلك السلوك وتنوع وتشابك المتغيرات المؤثرة به.

يجيء هذا العمل حلقة في سلسلة جهود، بدأت منذ زمن بعيد، عالجت القياس مفهوما ومنهاجا وأدوات. من هنا نعترف بأن هذا الجهد مكملا لما سبق، مفيدا لما هو قائم، وليس اختراعا أو اكتشافا بأي حال من الأحوال.

ومهما يكن من أمر، فإن هذه الطبعة الثانية من هذا الكتاب تشكل إضافة نوعية، ربما تغطي فراغا ما في المكتبة العربية. فهو حصيلة تدريس مواد تتعلق بالموضوع لأكثر من عشرين عاما، وثمرة الإطلاع على عدد ليس بقليل من المصادر والمراجع الحديثة في مجالات الإحصاء والقياس والتقويم والتقييم.

بدأ الكتاب بالفصل الأول يقدم وصفا لمفهوم القياس وخصائصه وأنواع المتغيرات والمقارنة بين القياس والتقويم والتقييم. ويعطي شرحا لمفهوم القياس الصفي وافتراضاته، موضحا أسباب عدم دقة القياس في العلوم السلوكية.

أما الفصل الثاني، فيقدم عرضا لآلية بناء وتطوير أدوات القياس عموما والإختبارات بوجه خاص، موضحا وصفا معقولا لجداول المواصفات، والمقارنة بين أنواع فقرات الإختبارات مدعما ذلك بالأمثلة الشارحة.

وقد استعرض الفصل الثالث علاقـة إعطـاء الـدرجات والتقيـيم المستمر، وبين طرق تصحيح وتقدير الدرجات للطلبة المفحوصين، مبينـا محاسن ومحددات كل طريقة. موضحا أن التقييم الأمثل لأداء الطالب هـو ما ينبع من تقييم أكثر من جانب من جوانب الفرد وبأكثر من أسلوب.

ويتعلق الفصل الرابع بتحليل أداء الأفراد على أداة القيـاس ككـل وعلـى الفقرات المكونة لها بـشكل مفـصل. تـم فيـه التعـرض للخـصائص السيكومترية لأداة القياس من خلال معامل الـصعوبة ومعامـل التمييـز وجاذبية المموهات مراعيا نوع الفقرات (الإنتقائية أو الصياغية).

وللحديث عن الخصائص السيكومترية للإختبار، يعـالج الفـصلان الخامس والسادس موضوعي الثبات والصدق. مدعما ذلك بالشرح المفصل للمفاهيم والتوضيح الكمي لها من خلال المسائل والأمثلة الكمية والنوعيـة ذات العلاقة.

ولتفسير مستوى أداء الفرد معياريا أو محكيا، يأتي الفصل الـسابع بعنوان المعايير. فقد تم التعرض لأنواع المعايير وطرق اشـتقاقها مـدعما ذلك بأمثلة من الواقع الميداني في قياس الذكاء غير اللفظي والقدرة علـى الرسم لدى فئات عمرية مختلفة من الأفراد. ويختتم الكتاب بالفصل الثـامن الذي يستعرض بشمولية واسعة لأدوات القياس فـي المجـالات الإنـسانية الثلاث المعرفية والوجدانية والمهارية. فقد قدم عرضـا مفـصلا لأنـواع أدوات قياس الذكاء لدى الإنسان واتجاهاته وميوله وقوائم شخصيته. كمـا ورد وصف أولي لأدوات قياس التحصيل وشرح مبسط لكيفية بنـاء تلـك الأدوات.

وفي الختام، فما كتاب أساسيات القياس في العلوم السلوكية إلا عمل يقدم جزءا من المعارف والحقائق للقارىء والطالب العربي في مراحل البكالوريوس والدبلوم والماجستير والدكتوراه. وهو عمل من صنع بشري ينشد الكمال وقد لا يبلغه. راجيا مسامحتي في حال العثور على خطأ لغوي أو منهجي.. وأكون أكثر سعادة إذا تم ارسالها لي لأغراض التصويب....... شاكرا لكم تعاونكم على أي من العنوانين:

mousa_nabhan1@yahoo.com or mousamn@agu.edu.bh

والله الموفق

موسى النبهان

الفصل الأول

مبادىء ومفاهيم أساسية للتقييم والقياس في العلوم السلوكية

Principles and Concepts of Assessment and Measurement in Behavioral Sciences

1

المخرجات التعليمية

بعد دراسة هذا الفصل يتوقع من الطالب أن يكون قادرا على أن

- يقارن بين طبيعة القياس في العلوم السلوكية والعلوم الطبيعية.
- يصنف المتغيرات حسب عدد من المعايير.
- يقارن بين مستويات القياس.
- يقارن بين الصفر المطلق والصفر الإفتراضي.
- يعرف مفهوم السمة وخصائصها.
- يعدد خصائص السمة.
- يبين الخصائص الرئيسية للقياس في العلوم السلوكية.
- يقارن بين مفاهيم القياس والتقويم والتقييم.
- يقارن بين التقويم البنائي والتقويم التجميعي.
- يقارن بين التقييم المعتمد على معيار وذلك المعتمد على محك.
- يعدد القرارات التي تتعلق بالأفراد المستهدفين من عملية التقييم.
- يعدد مصادر ضعف دقة القياس في العلوم السلوكية.

مقدمة

ترتبط كلمة **القياس** Measurement في التعامل اليومي بمعاني الوضوح والدقة. كما يرتبط الإنسان العادي بعدد من عمليات يقوم بها يوميا، كتلك المتعلقة بمسافة يمشيها في المزرعة أو داخل منزله، أو حجم كمية من الماء يشربها في اليوم، أو فترة زمنية يستغرقها للوصول إلى المطار الدولي، أو أقصى وزن يستطيع حمله والمشي به دون عناء كبير. وفي الغالب ما يتم التعبير عن كل من هذه العمليات أو الكميات بوحدات معينة على هيئة نقاط تعرف بالأمتار أو اللترات أو الساعات أو الكيلوغرامات على الترتيب. وجدير بالذكر أن تتصف مثل هذه القياسات المتعلقة بالطول أو الحجم أو الزمن أو الكتلة (الوزن) بخصائص الدقة والموضوعية. والتي من النادر ما كانت تبدو مشكلا من أي نوع (Magnusson, 1967).

أما عندما يتجه اهتمام الإنسان إلى التعامل مع قياس صفات مثل الدافعية، أو القدرة على التفكير المنطقي، أو الانطواء، أو القدرة على ضبط العواطف، أو مدى وعي الذات عند الشباب، فإن الصورة تبدو أكثر تعقيدا وتشعبا من ذي قبل. عندها لا بد من التعرض إلى مفهوم القياس وتعريفه. فبالرغم من الاتفاق على أهمية القياس، مازال الناس على اختلاف واضح في تعريفهم له. فبعضهم يقول بأن القياس وصف البيانات باستخدام الأرقام. وبعض آخر يقول بأن القياس عملية جمع معلومات عن خاصية معينة. ويعرف ستيفنز (Stevens, 1946) القياس بأنه عملية التعبير الكمي عن الخصائص والأحداث بناءا على قواعد وقوانين محددة.

في ضوء ذلك، أمكن تعريف القياس بصورة أكثر تفصيلا بأنه: تعيين فئة من الأرقام أو الرموز (أو العلامات)، مناظرة لفئة من الخصائص أو الأحداث، طبقا لقواعد محددة (علام، 2000). ويتضح من ذلك أن التعريف يشتمل على ثلاثة مجموعات هي:

(1) مجموعات الأرقام والرموز

(2) مجموعات الخصائص أو الأحداث أو السمات

(3) مجموعات القواعد أو القوانين

وتختلف طبيعة الخصائص والأحداث حسب المجال الذي تنتمي إليه هذه الخصائص. فقد يكون المجال علوما طبيعية، كالفيزياء. وعندها ستكون أحد الخواص موضع البحث هي سلوك المواد بناءا على شكل التوزيع الإلكتروني لذراتها. وقد يكون المجال علوما سلوكية تتعلق بمتغيرات عقلية ونفسية مثل الميول ونمط الشخصية والاستعداد ومستوى التحصيل والقلق وغيرها.

طبيعة القياس في العلوم السلوكية

يكاد يصلح تعريف القياس الوارد آنفا على كافة مجالات العلوم، إنسانية كانت أم طبيعية. ولكن الاختلاف الوحيد في ذلك، يكمن في مسألة القواعد والقوانين المحددة. ففي العلوم الطبيعية، يمكن إجراء عمليات القياس بطرق مباشرة، حيث أنها عمليات مقننة ومتفق عليها وتتعامل مع ظواهر ومواقف، تعريفها واضح، ووجودها ملموس في أكثر الأحيان، إضافة إلى وجود مقاييس و موازين معرفة جيدا، يمكن تفسير النتائج اعتمادا عليها بدرجة عالية من الثقة والموضوعية. ومثال ذلك قياس طول جسم أو وزنه أو درجة حرارته. هذا على الرغم من وجود ظواهر طبيعية معقدة نسبيا ويصعب قياسها مباشرة كما في المسائل الكهرومغناطيسية أو الظواهر الثيرموديناميكية وغيرها.

أما الظواهر في العلوم السلوكية عموما والعلوم النفسية والتربوية خصوصا، فتتميز بدرجة كبيرة من التعقيد وعدم الوضوح، الأمر الذي يجعل قياسها أكثر صعوبة وتعقيدا من قياس مثيلاتها الطبيعية. فقد يكون من السهولة دراسة التطورات التي تطرأ على ملامح بذرة تزرع في تربة معينة تحت شروط بيئية مضبوطة. في حين لن يكون ذلك سهلا عند دراسة

القدرة العقلية لفرد ما، عن طريق عزلها عن متغيرات بيئية أو حتى الإلمام بكافة المتغيرات المؤثرة في تلك القدرة. إضافة إلى صعوبة قياس تلك القدرة بأدوات مدرجة تدريجا معياريا (كالمتر أو الغرام). إذ أن قياس سمات شخصية الفرد أو درجة التوافق لديه، أو مستوى الذكاء العاطفي، أو درجة العدوانية عنده لن يكون دقيقا بمستوى دقة قياس درجة حرارة كمية من الماء، أو قياس سلوك المعادن عند تعرضها إلى حرارة عالية.

كل ذلك يعود إلى أنه عند قياس أي ظاهرة نفسية مثلا، لا بد من تعريفها أولاً، وتحديد الفئة العمرية المعنية في تلك الظاهرة ثانيا، والتعرف على مجمل الظروف التي ربما تؤثر في تلك الظاهرة وتغير في مقدارها، أو في دقة قياسها ثالثاً، إضافة إلى أن الظواهر النفسية والاجتماعية لا تتمتع بدرجة عالية من الثبات والاستقرار كما في الظواهر الطبيعية رابعا.

القياس وأنواع المتغيرات

عرفنا سابقا بأن قياس الظواهر أو السمات النفسية والتربوية يحتاج تعريفا دقيقا لكل منها في ضوء المتغيرات والعمليات المتعلقة بها، بهدف الحصول على قيم قابلة للتفسير أو التنبؤ. كذلك، لا بد أن يكون التعريف المطلوب إجرائيا Operational، يتم اشتقاقه من الأطر النظرية والبحوث والدراسات المتعلقة بالمتغيرات.

والمتغير Variable صفة أو خاصية من خواص شيء أو فرد، قد تأخذ أكثر من قيمة أو مستوى في الظروف والأوقات والحالات المختلفة (Glass and Hopkins, 1996). ومن الأمثلة على ذلك متغيرات: العمر، والجنس، واللون، والوزن، والطول، وغيرها. فالجنس يكون ذكراً أو أنثى، ويأخذ العمر قيماً متعددة، فقد يكون سنة، أو عشر سنوات، أو أكثر أو أقل. وتتنوع المتغيرات ومستوياتها بحسب طبيعة البحث وأهدافه، ومن أمثلة المتغيرات في المجالات التربوية والنفسية والاجتماعية التحصيل، والذكاء،

وطرق التدريس، ومستوى القلق، والاختيار الدراسي، ودرجة الرضا، وفعالية الأداء الإداري.

وعندما يتشابه مجموعة الأفراد في خاصية معينة، تكون هذه الخاصية بالنسبة لتلك المجموعة ثابتا Constant أي أنها ليست متغيراً. فلو كان مجال اهتمام باحث ما في استقصاء واقع تربوي في مدرسة للذكور، عندها يكون جنس الطلبة ثابتاً، ويعبر عنه بقيمة رقمية ثابتة، ولا يدخل الجنس كمتغير في هذا الاستقصاء. ولو كانت مجموعة مهنة الأب تتألف من مزارع، معلم، محامي، مهندس، طبيب، متقاعد، يمكن القول بأن مهنة الأب متغيرا، ويمكن تصنيفها في مستويات، يحمل كل منها رقماً.

أما إذا كان موضوع البحث يتضمن معلومات عن أولياء أمور التلاميذ في مدرسة لأبناء المتقاعدين، تكون مهنة الأب ثابتا (متقاعد) وليس متغيراً، وعندما تكون مؤهلات أولياء أمور الطلبة موضع بحث، وكانت المؤهلات متنوعة، عندها يكون المؤهل العلمي (جامعي، دبلوم،ثانوية، أمي) متغيرا. أما لو كان ذات المتغير موضع اهتمام في مدرسة معينة كمدرسة الجامعة النموذجية (هناك مدرسة نموذجية تتبع لمعظم الجامعات الرسمية في بعض البلدان) يصبح المؤهل العلمي لأولياء الأمور ثابتاً، خاصة إذا كان الالتحاق بتلك المدرسة مقصوراً على أبناء أعضاء الهيئة التدريسية في الجامعة (النبهان، 2009).

وبلغة أخرى، تعتبر الصفة متغيرا إذا كانت تتألف هذه الصفة في ظرف معين من أكثر من مستوى. وتعد تلك الصفة ثابتا عندما تقع في مستوى واحد. فالجنس، يكون متغيرا إذا تم التعامل مع مجموعة أفراد تتألف من مستويين (ذكور وإناث). ويكون الجنس ثابتا، إذا تم التعامل مع مجموعة أفراد تتألف من مستوى واحد فقط (الاناث).

تصنيف المتغيرات

تتنوع المتغيرات وتتحدد وفقاً لتصنيفات عدة، فعندما يتم التصنيف في ضوء موقع المتغير في الدراسة التجريبية يتبادر إلى الذهن نوعان من المتغيرات هما، المتغير المستقل Independent variable والمتغير التابع Dependent variable. ويعرف المتغير المستقل بالمؤثر أو المسبب، وهو المتغير الذي يتم التحكم به من قبل الباحث في التجارب الطبيعية أو النفسية أو الاجتماعية من أجل التعرف على تأثيراته في نتيجة التجربة. أما المتغير التابع فهو تلك النتيجة، أو المتغير الذي يقع عليه التأثير من المتغير المستقل. فمثلاً عند دراسة أثر اختلاف طريقة التدريس على تحصيل طلبة الصف السادس في اللغة العربية، فإن طريقة التدريس هي المتغير المستقل، بينما يكون التحصيل متغيراً تابعاً. إذ يستطيع الباحث أن يختار طرق التدريس فيدخل بعضها في الدراسة ويستثني البعض الآخر ولكنه يقيس ويسجل تحصيل الطلبة ولا يمكنه التدخل فيه زيادة أو نقصاناً. ويدعى المتغير المستقل أحياناً بالمتغير التجريبي، بينما يعرف المتغير التابع بالنتيجة أو المحك أو الظاهرة موضع الدراسة (Gay, Mills, and Airasian, 2009).

ويمكن تصنيف المتغيرات وفقاً لطبيعتها من حيث إمكانية التعبير عنها رقمياً إلى متغيرات كمية Quantitative variables ومتغيرات نوعية Qualitative variables. فالمتغير الكمي هو الذي يمكن التعبير عنه كمياً، أو يمكن تمييزه بالدرجة أو التكرار أو الكمية، كالتحصيل الذي يعبر عنه بدرجات الاختبارات (90 ، 88 ، 74، 56) أو الغياب الذي يعبر عنه بالتكرارات (6 مرات، 3 مرات، 11 مرة) خلال فصل دراسي معين،

وأطوال التلاميذ في الصف الثالث الأساسي (110 سم، 115 سم، 98 سم). ويمكن أن تكون قيم المتغيرات الكمية مرتبة في أولويات من مثل: الأول، أو الثاني، أو العاشر، أو العشرون (علام ، 2000).

أما المتغيرات النوعية فهي التي تصنّف الأفراد أو الأشياء حسب النوع، ولهذا فهي فاقدة صفة الترتيب، ومثال ذلك متغير الجنس الذي يصنّف الأفراد إلى ذكور أو إناث، ومتغير الفرع الدراسي الذي يصنف الطلبة إلى أدبي أو علمي أو مهني، وكذلك متغير التخصّص الدراسي الذي يصنف التخصصات في علوم طبية أو زراعية أو تربوية أو رياضية وهكذا. وتستخدم الأرقام في هذه المتغيرات لغايات التصنيف فقط، ولا تعكس مقادير كمية (Blommers and Hastings and peacock, 1975 ;Forsyth, 1977).

وتصنّف المتغيرات نظراً لطبيعتها من حيث القيم التي يمكن أن تأخذها إلى مستمرة أو متصلة Continuous ومتقطعة أو وثابة Discrete. ويكون المتغير مستمراً عندما يأخذ أية قيمة في مدى متصل وعلى مقياس معين، كالوزن، أو العمر، أو مستوى الذكاء، أو مستوى التحصيل. بينما يكون المتغير متقطعاً عندما يعطى قيماً بمثابة رموز ومسميات، كالجنس الذي يكون ذكراً أو أنثى فقط، ولا توجد بينهما قيمة أخرى، وعدد تلاميذ الصف أو عدد أفراد الأسرة وما شابه (Glass and Hopkins, 1996). ويمكن التمييز أيضاً بين المتغير المستمر والمتقطع بحيث أن قيم المتغير المستمر يمكن أن تأخذ قيماً كسرية مثل: (3.56)، (4.8)، (1.5) وهكذا، في حين لا يمكن للمتغير المتقطع أن يأخذ كسوراً ، ولكنه يأخذ أعداداً صحيحة فقط مثل:3 ، 8، 100، وهكذا (Gravetter and Wallnau, 1991).

هناك فرق بين القياس الحقيقي لقيم المتغير المتصل والقياس الذي يتم تسجيله فعلا نتيجة عملية القياس. بسبب احتمال وجود هامش معين من الخطأ، وهذا يعتمد على وحدة قياس المتغير. ويبين الجدول التالي

المقتبس من (Glass and Hopkins, 1996) القيم المعطاة لقياس قيمة متغير وحدي تلك القيم: الحد الأعلى = القيمة المعطاة + نصف وحدة القياس، والحد الأدنى = القيمة المعطاة – نصف وحدة القياس.

المتغير	وحدة القياس	القيمة المعطاة (كمثال)	حدا القيمة الفعليين	
			الأدنى	الأعلى
الوزن	الكيلوجرام	130	129.5	130.5
طول الملعب	المتر	66	65.5	66.5
طول الشخص	0.1 سم	170.5	170.45	170.55
زمن الاستجابة	0.01 ثانية	0.49	0.485	0.495
مستوى التحصيل	المئين	% 78	% 77.5	% 78.5

مستويات (موازين) القياس

يعد التعرف على مستويات القياس Scales of Measurement من الأساسيات التي ينبغي للباحث في العلوم السلوكية التعرف عليها، كي يستخدم الطرق الإحصائية المناسبة لتحليل البيانات الخاصة بتجاربه وبحوثه استخداما مناسبا. وهذا يقود إلى أن ينتهي البحث بنتائج صادقة وقرارات أكثر فاعلية(Glass and Hopkins, 1996).

تصنف مستويات المقاييس في أربعة أنواع هي: (1) المقاييس الاسمية Nominal Scales، (2) المقاييس الرتبية Ordinal Scales، (3) المقاييس الفئوية Interval Scales، (4) المقاييس النسبية Ratio Scale. والجدير

بالذكر أنه يتم عادة تصنيف المتغيرات وفقاً لهذه المقاييس. فهي بمجموعها ذات صفة تراكمية، بمعنى أن مستويات القياس متضمنة فيما بينها. فمثلاً يتضمن المقياس الرتبي خصائص المقياس الأسمى، وكذلك يحوي المستوى الفئوي خصائص كل من المستويين الرتبي والاسمي، كما يتضمن المستوى النسبي كافة خصائص المستويات الأخرى (Glassnapp and Poggio, 1985; Ferguson, 1966).

1 - مستوى القياس الاسمي

وتمثل المرتبة الأدنى للقياس، وتستخدم في معظم الأحوال مع المتغيرات النوعية، حيث يتولى المقياس الاسمي تصنيف الأفراد أو الأشياء في عدة مجموعات وفقاً لبعض الخصائص النوعية كتوزيعهم حسب جنسهم (ذكور وإناث)، أو حسب مناطقهم السكنية (جنوب، شرق، شمال، غرب)، أو حسب لون العيون (أزرق، أخضر، عسلي، أسود)، أو حسب الرقم الوطني لكل شخص، أو حسب لون البشرة، أو في ضوء الأرقام المكتوبة على صدور اللاعبين، والرقم الجامعي(Hieman, 1992; Kagan and McCall, 1988).

لا تعمل هذه المقاييس أكثر من تصنيف الأشياء من أجل التمييز بينها اعتماداً على افتراض أن الأفراد يختلفون في صفة ما (الجنس، اللون، الذكاء، الرتبة العسكرية، الأسماء). ولتسهيل التعامل مع هذه المتغيرات، وخزنها، ومن ثم عرضها في الحاسوب، بات من الضروري تكميمها (ترميزها)، أو التعبير عنها رقمياً (Glass and Hopkins, 1996). ذلك بإعطاء كل صفة رقماً أو مقداراً للتعرف عليه وتصنيفه فقط. وهذا الرقم لا يفيد في أكثر من التسمية أو التصنيف. إذ أن الأرقام في هذا المستوى أشبه بالأسماء والألقاب، فهي لا تتضمن معنى للأفضلية (الأكبر والأصغر). فمثلاً إذا أعطى للذكر الرقم (1) وللأنثى الرقم (2)، فهذا لا يعني أن الذكر أقل من الأنثى أو أن الأنثى أفضل من الذكر. وكما هو الحال بالنسبة لأرقام

اللاعبين (1، 2، 3، ... 13) أو الأرقام الجامعية، فهي تصنيفية تستخدم للتسمية بدلاً من استخدام الأسماء الفعلية للرياضيين أو للطلبة.

مستوى القياس الرتبي

يُعتبر هذا النوع من المقاييس تالياً من حيث المستوى للمقاييس الإسمية، فهو أعلى منها، فبالإضافة إلى تصنيف الأفراد والأشياء في مجموعات متمايزة، فهو يرتب الأفراد والأشياء تصاعدياً أو تنازلياً في صفة أو خاصية معينة. وعندما تعطى الأرقام للأشياء والأفراد وفقاً لهذا المقياس، فإن تلك الأرقام لا تمثل كميات معينة. بل تشير إلى ترتيب مقدار امتلاك الصفة. كذلك، فإن المسافات الفاصلة بين رتبة وأخرى لا يشترط أن تكون متساوية، فمثلاً عند ترتيب خمسة طلاب حسب درجة تعاونهم، ويعطى رقم (1) لأكثرهم والرقم (5) لأقلهم. فإن الفرق بين درجتي تعاون الثاني والثالث لا يساوي الفرق بين درجتي تعاون الثالث والرابع. كذلك لا يشترط أن يكون تعاون الطالب الأول مساوياً لخمسة أضعاف تعاون الطالب الخامس. فالمقياس الرتبي لا يعطي صورة واضحة عن حجم الفروق الموجودة بين الأفراد المتجاورين في أية مجموعة (Glass and Hopkins, 1996 ; Glassnap and Poggio, 1985).

ومن الأمثلة على المقياس الرتبي، درجات المدرسين حسب الكفاءة التدريسية ، ورتب التلاميذ حسب نشاطهم في الصف، وترتيب الفرق الرياضية في لعبة معينة حسب مستوى الجدارة والكفاءة في تلك اللعبة، وكذلك رتب مجموعة من الضباط (Hinkle, Wiersma, and Jurs, 1979). فعند القول بأن رتب الضباط تقاس على المستوى الرتبي، نفهم بأن الضباط يختلفون حسب الرتبة العسكرية (وهذه خاصية للمقياس الإسمي)، ثم نعرف بأن الرتب تتفاوت فيما بينها، فرتبة العميد أعلى من رتبة الرائد، ورتبة الرائد تفوق رتبة الملازم وهكذا (وهذه هي خصائص المقياس الرتبي). ولكن ليس الفرق بين رتبة العميد ورتبة العقيد كالفرق بين رتبة

الرائد ورتبة الملازم. وكمثال آخر، ليس الفرق بين علامتي الطـالبين ذي الترتيب الأول و ذي الترتيب الثاني كالفرق بين علامتي الطالبين الثالث والرابع. وقد تكون المعدلات الفعلية لخمس من الطلبة في صف معين هي 98، 90، 87، 60، وبنفس الوقت ربما تكون رتبهم في الصف هي (1، 2، 3، 4). الأمر، الذي يفيد بأن الفرق بين الرتب لا يشير إلى الفرق بين مدلولات هذه الرتب.

مستوى القياس الفئوي

يُعتبر هذا النوع من المقاييس أعلى مستوى من المقياسين السابقين، وهو يمتلك خاصية الفواصل أو المسافات المتساوية التي تفصل بين كل درجتين متجاورتين. هذا يشير إلى إمكانية إجراء عمليات حسابية لإيجاد الفروق، فعندما تكون درجات أربع طلاب هي 55، 60، 70، 75، فإنه يمكن القول بأن الفرق بين (55) و (60) هو (5) ويساوي الفرق بين (75) و (70). كما يجوز القول بأن الفرق بين العلامتين (75) و (60) يساوي ثلاثة أمثال الفرق بين العلامتين (60) و (65). ولكن لا يمكن الاعتقاد بأن العلامة (70) تساوي ضعف العلامة (35) (Glass and Hopkins, 1996).

والجدير بالذكر، بأن الصفر في هذا المستوى افتراضي وليس حقيقياً. بمعنى أن الصفر هو قيمة كباقي القيم، وهو لا يعني غياب الصفة أو انعدامها. فمثلاً لا تعني درجة الصفر المئوي عدم وجود حرارة، بل إن الجسم الذي درجة حرارته تساوي الصفر يحوي حرارة أكثر من الجسم في الدرجات السالبة (‎-1 ، ‎-4 وهكذا ...). وكذلك لا يمكن الاعتقاد بأن العلامة (صفر) لطالب ما في امتحان معين تعني أنه لا يعرف شيئاً (أو أن تحصيله صفراً). فالصفر هنا افتراضي وليس حقيقياً، الأمر الذي لا يجوز في هذا المستوى إجراء عمليات حسابية كالجمع أو الضرب. فمثلاً لا يمكن القول بأن (10 درجات مئوية + 10 درجات مئوية = 20 درجة مئوية) أي أنه إذا خلطنا مقداراً من الماء درجة حرارته 10 درجات مئوية

مع مقدار مشابه من الماء وبنفس الدرجة، لن تكون درجة حرارة المـزيج عشرين درجة. وكذلك لا يجوز القول بأن العلامة (80) هي ضعف العلامة (40)، ولا يمكن القول بأن عبقرياً درجة ذكائه (150) يقابل (يعادل) ثلاثـة أشخاص متخلفين عقلياً درجة ذكاء كل منهم (50). أي أنه ليس من الممكن أو الصحيح القول بأن حاصل قسمة درجة ذكاء عبقـري (150) علـى (3) يكون الجواب درجة ذكاء شـخص غبـي (50) (Witte, 1993; Howell, 1999).

ومن الأمثلة على المتغيرات التي تقاس على المقياس الفئوي معدلات الذكاء ودرجات الحرارة المئوية أو الفهرنهيتية ودرجـات الطالـب علـى الاختبارات التحصيلية، وأرقام السنوات على اعتبار أن السنة (1)، وهـي موعد ميلاد السيد المسيح لا تمثل صفراً حقيقيا، بل هي نقطة مرجعية تمثل بداية التقويم الميلادي.

مستوى القياس النسبي

تُعتبر المقاييس النسبية أعلى مستويات القياس، وتتميز بخصائص جميع المقاييس السابقة، إضافة إلى وجود الصفر المطلق أو الحقيقي (والذي يعني غياب الصفة). فعندما نقول إن فلاناً يملك صفراً من النقود فهذا يعني أنه لا يملك نقوداً أبداً. وكذلك يمكن القول بأن المسافة التي تبلغ (100) كيلـومترا تساوي ضعف تلك التي تساوي (50) كيلومترا، أو أربعة أمثال المسافة التي تبلغ (25) كيلومترا وهكذا. ومن الأمثلة علـى المتغيـرات التـي تتمتـع بخصائص مستوى القياس النسبي كالدخل والوزن والطول ودرجة الحرارة المطلقة. وجدير بالذكر أنه يمكن في حالة المقياس النسبي استخدام جميـع العمليات الحسابية من جمع وطرح وضرب وقسمة (Spatz and Johnston, 1989).

وبالختام، يمكن القول بأن القياس الاسمي يصنف المتغيـرات بـدون إعطائها رتبة، بينما يصنف المقياس الرتبي المتغيرات بدرجات ورتب. أما

المقياس الفئوي، فيصنّف المتغيرات برتب ويوفر وحدات متساوية في المقياس، وبالإضافة لكل الصفات الواردة يوفر المقياس النسبي نقطة الصفر الحقيقي.

ولا بد من الإشارة هنا إلى أن المستوى الرتبي أكثر فائدة من المستوى الفئوي أو النسبي في حالات شائعة، كمثال قولنا بأن طول هذا الطفل هو أكبر من (80 %) من أبناء جيله (رتبي) فضلاً عن القول بأن طوله يساوي (1.75) مترا (نسبي). أو عندما نقول إن معدله في الرياضيات يفوق (90 %) من أبناء صفه، يكون ذلك ربما أفضل وأكثر إفادة ودلالة من القول بأن معدله يساوي (95)، وخاصة لدى المهتمين بموضوع مقارنة الأداء بين الأفراد. والجدول التالي يبين وصفا للخصائص الرئيسة لكل مستوى من مستويات القياس:

مستوى القياس	الخاصية			
	الصفر الحقيقي	المسافات متساوية	المقدار أو الكمية	الفروق
الإسمي				xxxxxxxx
الرتبي			xxxxxxxxx	xxxxxxxx
الفئوي		xxxxxxxxx	xxxxxxxx	xxxxxxxx
النسبي	Xxxxxx	xxxxxxxxx	xxxxxxxxx	xxxxxxxx

ويبين المخطط التالي وصفا ملخصا لمستويات (موازين) القياس مبينا خصائص كل منها والعمليات الكمية الممكن إجراؤها مع الأمثلة:

مستوى القياس	خصائص القياس	الأمثلة	العمليات الممكنة
النسبي	يوجد صفر حقيقي. وهذا يعني غياب الصفة. وتمثل الأرقام كميات متساوية عن موقع الصفر المطلق.	العمر، الدخل، درجة الحرارة المطلقة، السرعة، الزمن، حجم الصف، عدد أفراد الأسرة.	يمكن إجراء كافة العمليات الحسابية وجميع مقاييس النزعة المركزية والتشتت. وللنسبة معنى.
الفئوي	الفروق بين القيم تساوي الفروق بين قيم المتغير أو قيم السمة التي تخضع للقياس. الصفر افتراضي (غير مطلق)، ولا يعني هذا الصفر غياب أو انعدام الصفة.	درجة الحرارة المئوية، التقويم السنوي، الدرجات على الاختبارات العقلية، معدلات الذكاء.	يمكن إجراء عمليات حسابية على قيم المتغير الواحد. وليس للنسبة أي معنى.
الرتبي	تمثل الأرقام رتبا للأفراد حسب درجة امتلاكهم للسمة موضع البحث. تمثل الأرقام صفة كمية وتشير إلى مفهوم الأفضلية (الأصغر والأكبر). ليس للفروق بين الرتب أي معنى أو أهمية.	المستوى الاجتماعي والاقتصادي، رتب الطلبة في مستوى تحصيلهم الدراسي، الرتبة الاكاديمية، الرتبة العسكرية.	لا يمكن إجراء عمليات حسابية، يمكن التعامل مع المنوال والوسيط فقط. ويتم التعبير عن القيم بالمئينات فقط.

الاسمي	تستخدم الأرقام لأغراض التصنيف، ولا تمثل أي صفة كمية أو ترتيب معين، ولا يخضع استخدام الارقام في التصنيف إلى قوانين أو قواعد.	الجنس، الجنسية، نمط الشخصية، اللغة، الديانة، الرقم الجامعي.	لا يمكن إجراء أي عمليات حسابية. ويمكن التعامل مع المنوال فقط.

هرمية مستويات القياس

عرفنا أن المستوى النسبي يتضمن خصائص المستوى الفئوي وكذلك المستويين الرتبي والاسمي. بمعنى أن جميع المتغيرات تقع في المستوى الاسمي بفضل وجود الفروق والاختلافات في امتلاك الصفة التي يشير إليها المستوى. لكن لكل متغير مستوى أعلى يصنف فيه المتغير. فعليه، يمكن القول بأن متغير الجنس مثلا يقع في المستوى الاسمي وهو أعلى مستوى يمكن أن يصنف فيه ذلك المتغير. أما الدخل مثلاً، فيقع في كل المستويات ولكن أعلى مستوى يمكن أن يصنف فيه هو المستوى النسبي. ويعرض المخطط التالي فكرة هرمية مستويات القياس. إذ يشير الخط الغامق إلى المتغير في أعلى مستوى يمكن أن يصنف فيه.

المستوى النسبي	**الوزن، العمر، الطول، الدخل.**
المستوى الفئوي	**درجة الحرارة المئوية، العلامات على اختبار عقلي، العلامات على اختبار تحصيلي، معدل الذكاء، الوزن، العمر، الطول، الدخل**
المستوى الرتبي	**الرتبة الأكاديمية، الرتبة العسكرية، درجة الحرارة المئوية، العلامات العقلية، المستوى التعليمي، الوزن، العمر، الطول، الدخل، العلامات على اختبار تحصيلي، معدل الذكاء.**
المستوى الاسمي	**الجنس، الجنسية، اللغة، الديانة، الرتبة الأكاديمية، الرتبة العسكرية، درجة الحرارة المئوية، العلامات العقلية، المستوى التعليمي، الوزن، العمر، الطول، الدخل، العلامات على اختبار تحصيلي، معدل الذكاء.**

تصنيفات أخرى للمتغيرات

يمكن تصنيف المتغيرات حسب تضمينها في البحث أو التجربة. فهناك متغيرات مضبوطة Controlled يلغي الباحث أثرها في التجربة لتقليل مصادر الخطأ، وأخرى معدلة Moderator والتي تعتبر بمثابة متغيرات مستقلة ثانوية تقع تحت سيطرة الباحث في تجربته. وهناك أيضاً متغيرات دخيلة Intervening وهي نوع من المتغيرات المستقلة التي لابد من ضبطها أو منع تأثيرها في الدراسة باتباع أسلوب معين من أساليب ضبط المتغيرات. والجدير بالذكر، فقد يرد تفصيل مناسب عن شرح هذه المتغيرات في مواضيع تتدرج تحت موضوعات مناهج وتصميم البحوث العلمية (Gay, Mills, and Airasian, 2009).

ولا شك بأن المتغير الواحد يمكن تصنيفه في أكثر من فئة تصنيفية. فمتغير كالجنس مثلاً يُقاس في المستوى الاسمي من ناحية، ويصنف على أنه متغير نوعي ومتقطع من ناحية أخرى. وكذلك يعمل على كونه متغيراً مستقلاً، أو يدعى أحياناً متغيراً عضوياً Organismic.

كما لابد من التنويه بأن التحصيل مثلاً، يمكن أن يكون متغيرا مستقلاً أو تابعاً، ويمكن أن يكون كمياً متصلاً وفئوياً عندما يعبر عنه بالمعدل التراكمي، أو قد يتم تحويله إلى متغير إسمي (راسب، ناجح)، أو إلى درجات من نوع الأول والثاني (رتبي).

من ناحية أخرى، يصنف المتغير الكمي في كونه فئويا أو نسبيا أو رتبيا. بينما يصنف المتغير النوعي في كونه متغيرا إسميا. أما بالنسبة للمتغير المستمر، فيمكن أن يكون فئويا أو نسبيا، في حين يصنف المتغير المتقطع في كونه رتبيا أو إسميا. ويتم تلخيص ذلك كما في الشكل التالي:

علاقة المتغير (كمي ونوعي) و (متصل ومتقطع) بمستويات القياس

المتغير				
كمي			نوعي	
متصل		متقطع	متصل	متقطع
نسبي	فئوي	رتبي	لا يوجد	إسمي

السمة : مفهومها، خصائصها

السمة trait عبارة عن تجمع من السلوك المترابط الذي يحتمل أن يحدث معا. وهي صفة مركبة غير منفردة، كما أنها مفهوم مجرد غير ملموس. وتكون عامة كسمة "الميل نحو العمل التعاوني" ، أو سمة محدودة كسمة "تذوق الطعام" (علام، 2000).

ويستدل عادة على السمة من خلال أنماط السلوك الملاحظ، ولا يمكن قياسها مباشرة. فمثلا عند قياس سمة عقلية كالذكاء، لا يمكن رؤية الذكاء أو ملاحظته مباشرة، بل يمكن ملاحظته من خلال مواقف مختلفة تتطلب الذكاء، كما في القدرة على التعامل مع الأرقام والأعداد، أو القدرة على فهم واقع معين من خلال عدد من المعطيات. ولكي يتم الاستفادة من السمات في وصف السلوك، فإنه لا بد من توفر عدد من الشروط الأساسية **Kelly, 1974; Nunnally, 1978)** التالية:

(1) أن تمثل السمة صفة عقلية أو وجدانية يتفاوت الأفراد في امتلاكها (يتم أحيانا استخدام السمة والقدرة كمترادفين مع أن القدرة تشير إلى خاصية عقلية أو جسمية، أما السمة فتشير إلى خاصية نفسية أو وجدانية).

(2) أن تكون السمة قابلة للتعريف الإجرائي من حيث إمكانية أن تتضمن الجوانب المهمة التي نريد التركيز عليها من حيث الفائدة العملية وإمكانية قياسها.

(3) أن تكون السمة ثابتة نسبيا بمرور الزمن مهما اختلفت المواقف. وهنا لا بد من الإشارة إلى أن السمات العقلية (الذكاء) عموما تمتاز بثبات نسبي أكثر من السمات الوجدانية أو الشخصية (القلق).

(4) وعند قياس السمة نفترض أن هناك متصلا من السلوك، ونقصد بالمتصل خط مستقيم يتكون من عدد لانهائي من النقاط الممكنة تحدد مواقع مختلفة للسمة موضع البحث. إذ غالبا ما تمثل السمات العقلية متصلا وليس تقديرا ثنائيا. فقد يصنف الأفراد على سمة الذكاء في مستويات مختلفة متدرجة يتمايزون فيها بحسب درجاتهم في هذه السمة، ولا يمكن تصنيفهم ثنائيا (ذكي، غبي). ويمكن تدريج متصل السمة بأسلوب يصبح معه بمثابة ميزان **scale** يعتمد على طبيعة السمة المراد قياسها.

(5) تتشكل السمات بفعل عوامل البيئة والوراثة، وأن جميع جوانب النمو (الجسمي والعقلي والوجداني) متداخلة ومتكاملة وذات طبيعة تطورية، كما أن هناك تغير في النمو يختلف بنسب متفاوتة حسب المرحلة العمرية، لذلك فإن التغير في النمو يختلف من سمة إلى أخرى، ومن شخص إلى آخر لذات السمة (عودة، 2000). الأمر، الذي يجعل فهم المقيم لشخصية الفرد المستهدف بعملية التقييم أمرا مهما في الوصول إلى عمل ذي مستوى نوعي. هذا يشير إلى أن إخفاق تقييم شخص ما ربما يعود إلى أن التقييم كان قد ركز على جانب معين (عقلي فقط) ولم يأخذ بعين الاعتبار الجوانب الأخرى (الوجدانية أو المهارية).

خصائص القياس في العلوم السلوكية

يتنوع ويتفاوت مدى اهتمام المشتغلين في التربية أو علم النفس بالسمات التي يمكن التعامل معها. وفي الواقع تحتل السمات التي تتعلق بالعمليات العقلية كالتحصيل والاتجاهات أولوية بينة. ويتصف قياس مثل هذه السمات بالخصائص الرئيسة التالية:

(1)**غير مباشر Indirect**، أي أنه لا يتم قياس السمة أو مقدار ما يمتلكه الفرد من تلك السمة بشكل مباشر، بل من خلال الأداء على مواقف لها علاقة بالسمة ذاتها. فالتربوي يستطيع التعرف على السمة من خلال السلوك الظاهري القابل للملاحظة وليس من خلال السمة ذاتها. ونحن لا نستطيع قياس القدرة على الاستدلال العلمي مباشرة، بل من خلال مستوى الأداء على عدد من المهمات والقدرات والمواقف والفقرات التي يتألف منها اختبار ورقة–قلم وتشكل بمجموعها سمة الاستدلال العلمي.

وإن تطوير أداة (إستبانة) لقياس اتجاهات طلبة الصف العاشر نحو تدريس مادة الحاسوب يعد مثالا على كون القياس النفسي أو التربوي قياسا غير مباشر.

(2)**غير تام Incomplete**، أي أنه لا يتم قياس السمة عن طريق التحقق من مستوى الأداء على كافة المثيرات والفقرات أو الأسئلة التي تمثل المجال السلوكي الذي تتكون منه السمة، بل يتم عادة اختيار عينة من تلك المثيرات يتكون منها اختبار يعتقد بأنه يقيس السمة. واختبار الذكاء غير اللفظي مثلا يمثل عينة من الفقرات التي تشكل بمجموعها اختبار الذكاء الذي يستخدم لأغراض متعددة. وكذلك يعتبر اختبار الرياضيات للصف الأول الثانوي بمثابة عينة من عدد من الفقرات التي تمثل محتوى الرياضيات للصف المذكور. في ضوء ذلك يمكن الإستنتاج بأن قياس سمتي الذكاء والتحصيل

وغيرها لا يمكن أن يكون إلا باستخدام عينات من الفقرات والأسئلة. وهذا إنما يشير إلى القياس النفسي أو التربوي غير تام.

من ناحية ثانية، يكون القياس الطبيعي تاما، وذلك لأن العينة التي يتم اختيارها في القياس تمثل تماما المجال الذي تسحب منه. إذ أن عينة الدم التي يتم فحصها لمعرفة خصائصه البيولوجية تغني تماما عن التعامل مع كامل الدم.

(3)**نسبي Relative**، إذ أنه لا يمكن تفسير درجة الفرد على الاختبار النفسي أو التربوي دون الرجوع إلى مستوى أداء المجموعة التي ينتمي إليها ذلك الفرد، أو بمقارنة مستوى أداء الفرد بمستوى أو بمحك **Criterion** أو بمعيار **Norm**، أو بحد أدنى معين **Benchmark**. فقد تكون الدرجة التي يحصل عليها الطالب منخفضة أو قد تكون عالية في ضوء مقارنة تلك الدرجة مع متوسط درجات أفراد المجموعة التي ينتمي إليها هذا الطالب. فإذا كانت درجة الطالب على امتحان (60) مثلا، فستكون تلك الدرجة عالية جدا إذا كان متوسط درجات المجموعة التي ينتمي إليها الطالب (50) وانحراف معياري (2)، ذلك لأن تلك الدرجة تفوق متوسط أداء المجموعة بواقع خمسة انحرافات معيارية. ولكن تكون الدرجة ذاتها منخفضة نسبيا، إذا كان متوسط المجموعة (50) وبانحراف معياري (10).

من ناحية ثانية، من الممكن أن يتم تقييم الدرجة بالمقارنة مع مستوى أو محك محدد مسبقا أو متعارف عليه ضمن معطيات معينة. فقد تعد القدرة على قفز جدار ارتفاعه 50 سنتيمترا من قبل طفل عمره 8 سنوات إنجازا كبيرا، بينما يعتبر عملا بسيطا إذا قام به شاب عمره 18 سنة.

وجدير بالذكر أن صعوبة القياس النفسي وضعف دقته النسبية تبدو واضحة بسبب صعوبة الحصول على الصفر الحقيقي (المطلق) والاعتماد

على افتراض أن الصفر في قياس السمات النفسية والتربوية افتراضي، وهو ما يمثل الحد الأدنى لمستوى الأداء. إذ أن وجود الصفر المطلق في الصفة المقاسة يعني أن تلك الصفة يمكن قياسها على المستوى النسبي ويعبر عنها بدقة كبيرة.

فعندما يتم قياس متغير على المستوى النسبي تتحقق دقة كبيرة في القياس، من مثل القول بأن المسافة (100 متر) أكبر من المسافة (50 متر) وتساوي ضعفها، وتساوي أربعة أمثال المسافة (25) مترا. في حين لا تتوفر الدقة بنفس الدرجة عند قياس متغير على المستوى الفئوي الذي لا تتعدى الدقة فيه القول بأن الفرق بين درجتي الذكاء (100) و (80) يساوي الفرق بين (60) و (40). وكذلك بالنسبة للقياس على المستوى الرتبي الذي لا تتعدى درجة الدقة فيه القول بأن معدل الذكاء (100) أكبر من معدل الذكاء (50) فقط. ناهيكم عن أن الدقة في المستوى الإسمي لا تتعدى القول بأن درجات الأفراد على مقياس الذكاء مختلفة فقط.

القياس والتقويم والتقييم

يحدث خلط والتباس في غالب الأحيان بين مفاهيم **القياس** Measurement **والتقويم** Evaluation **والتقييم** Assessment. فالقياس وصف كمي لمقدار السمة التي يمتلكها الفرد، ولا يرقى ذلك إلى إصدار أحكام حول تلك السمة. فإذا حصل شخص في اختبار ذكاء على الدرجة (120)، فإن هذا يعد قياسا، ولكن الحكم على تلك الدرجة وتفسير المستوى الذي تدل عليه يعد تقويما. ولأن التقويم عملية إصدار حكم أو وصف كمي أو نوعي للدرجة أو لمستوى الأداء، فإنه يتوقع أن يفيد في اتخاذ قرار معين بشأن الفرد الممتحن أو بشأن التوجيه نحو اتخاذ إجراء مناسب حيال ذلك.

والمعروف أن التقويم يشمل القياس. وأن القياس أساس مهم لتسهيل القيام بالتقويم وجعل احتمالات الخطأ فيه أقل ما يمكن، حيث أن العلاقة

بينهما واضحة ولكن الفصل بينهما غير ممكن. فإذا كان القياس تعبير رقمي عن السمة أو الأحداث بناءا على قواعد وقوانين معينة، فإن التقويم عملية منظمة لجمع وتحليل المعلومات لتحديد مدى تحقق الأهداف المرسومة لفصل دراسي أو حصة دراسية أو مشروع تدريبي أو إرشادي معين.

ويتركز الهدف الرئيسي لتقويم برنامج تدريبي أو تربوي في التزويد بمعلومات لاتخاذ القرارات بشأن هذا البرنامج، ويتم ذلك بوجود عدد من البدائل الممكنة. مبينا أن يتوقع من التقويم أن يكون قرارا دافعا نحو تطوير البرنامج، لا أن يتركز القرار حول الاستمرار في البرنامج أو التوقف عنه فقط.

وحتى يكون التقويم نشاطا فاعلا، يجب أن يكون مراعيا للتوقيت المناسب لمتخذ القرار خشية أن يكون القرار عتيقا وغير مواكب أو ملائم. ومن المفروض أن يكون المقومون على دراية بالأمور التالية التي تساعد في أن يتمتع قرار التقويم بدرجة أكبر من الصدق والفعالية:

(1) هناك أهداف طويلة المدى وأخرى قصيرة المدى ينوي البرنامج المنوي تقويمه تحقيقها،

(2) من الحيوية أن ينتبه المقومون لتأثير أية أحداث غير مقصودة تؤثر على سير البرنامج موضع البحث،

(3) حتى يتحلى قرار التقويم بالدقة والفاعلية والموضوعية، فإن من المفروض أن تتمتع أدوات التقويم بدرجة عالية من الصدق والثبات والموضوعية.

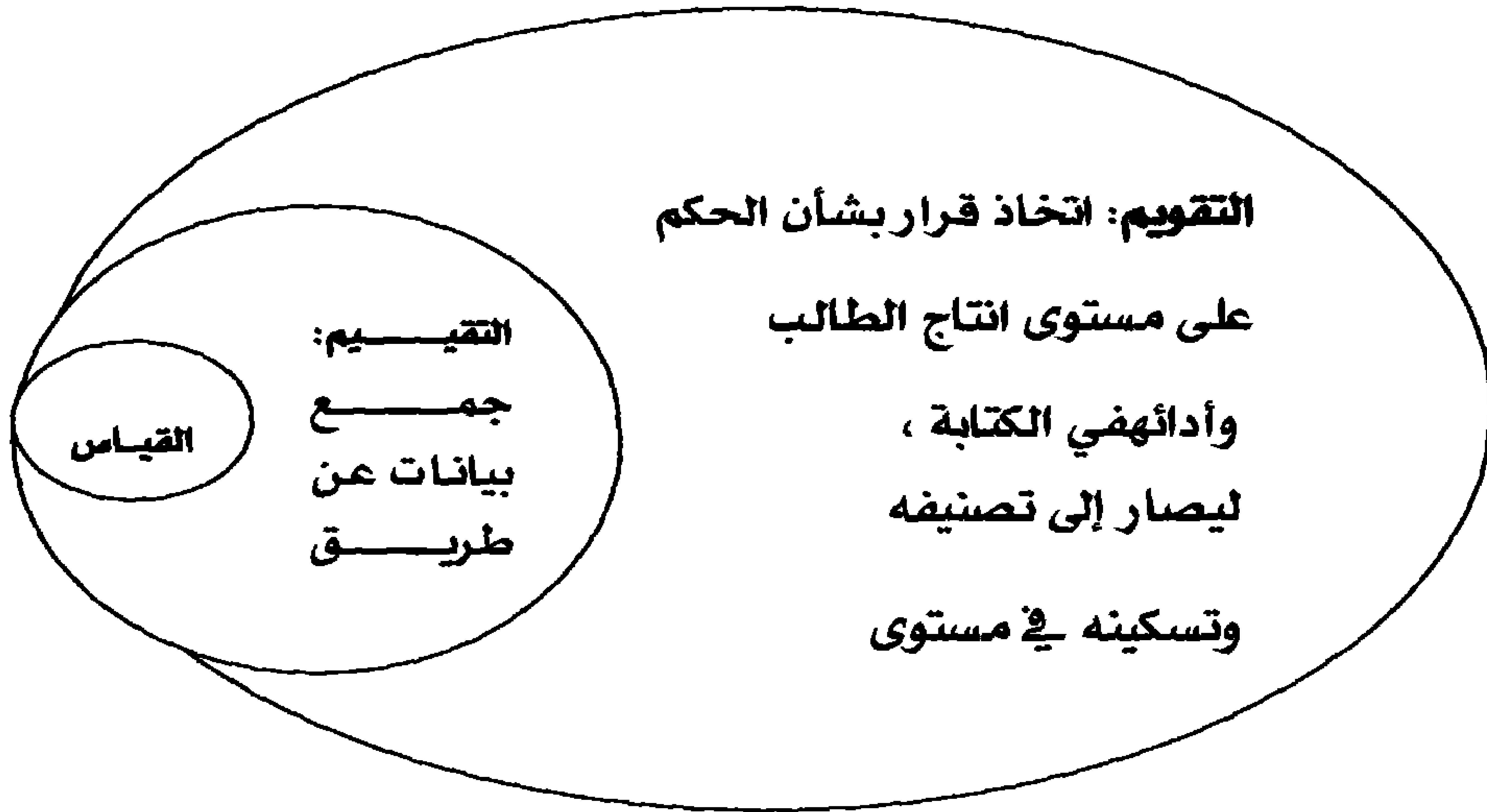
أما التقييم، فقد شاع استخدامه مؤخرا في مجال قياس الشخصية وفي الميادين الإكلينيكية وفي وصف أداء Performance وإنتاج Product الفرد في المعمل أو الصف الدراسي أو البرنامج التدريبي في ضوء عدد من المحكات Criteria كالامتحانات التحريرية examinations الطويلة

منها والقصيرة quizzes والملاحظات observations والمقابلات interviews، والعروض الشفوية أو العملية oral or practical demonstrations، والتصاميم designs والنتاجات products و ملفات الانجاز portfolios وغيرها.

وبالرغم من أن مفهومي التقويم والتقييم يستخدمان كمترادفين في أغلب الأحيان، إلا أن التقييم أكثر خصوصية من التقويم وأكثر اتساعا وشمولا من القياس. وقد عرف نيتكو (Nitko, 2001) التقييم بأنه عملية جمع ووصف وتكميم البيانات المتعلقة بمستوى أداء معين بقصد استخدامها في اتخاذ قرارات معينة، فقد تكون القرارات متعلقة بالتوصل إلى تصورات وتفسيرات أو اختبار فرضيات تتعلق بسلوك فرد أو استعداده لممارسة نشاط تعليمي، أو حتى ممارسة مهنة معينة.

مثال توضيحي عن علاقة التقويم بالتقييم والقياس

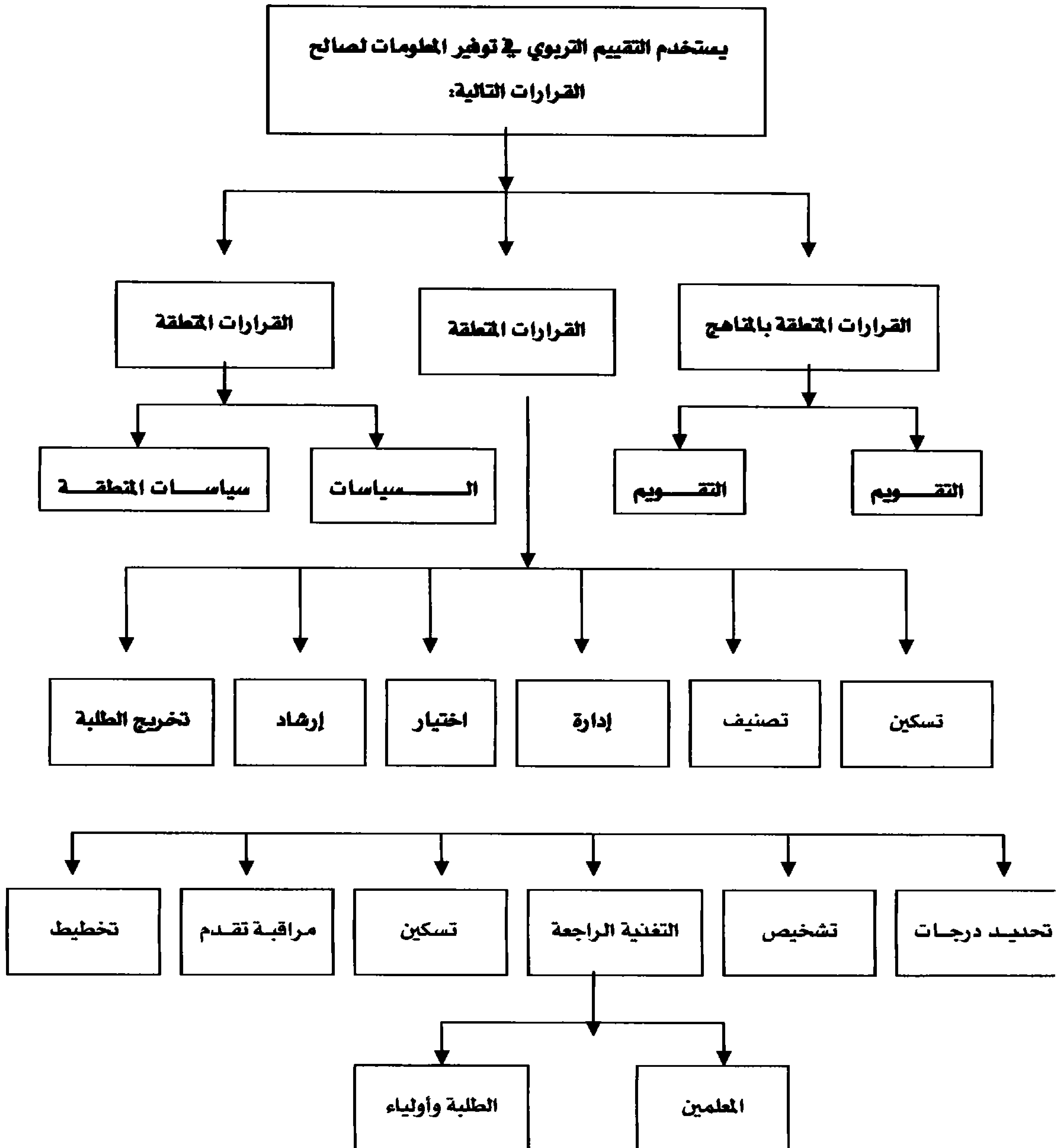
إذا كان **التقويم** عملية اتخاذ قرار ذي قيمة يتعلق بأداء أو نتاج طالب ما، بقصد الحكم على مستوى قدرته في الكتابة، ليصار إلى تصنيفه وتسكينه في مستوى دراسي معين، فإن **التقييم** يتضمن تحديد قدرته على الكتابة من خلال جمع البيانات عن طريق مراجعة كتاباته ومقارنتها مع أعمال أقرانه، ومعرفة مدى تحقيق أعماله لمعايير الكتابة الصحيحة. كما قد تتم عملية التقييم بناءا على نتيجة الطالب في أدائه على اختبار كتابي مقنن لهذا الغرض (وهذا يعد **قياسا**). إن هذا كله يساعد حتما في التحضير لاتخاذ قرار بشأن تصنيف الطالب ومن ثم التوجيه بوضعه في المستوى الدراسي المناسب. والشكل التالي يبين العلاقة بين التقويم والتقييم والقياس بناءا على المثال أعلاه:



يعتمد التقويم أو ربما لا يعتمد على القياس ونتائج الاختبارات. فقد يتم تقويم الطلبة اعتمادا على الملاحظات المنظمة والوصف النوعي دون إجراء قياس. وحتى لو توفرت المعلومات الموضوعية عن الطالب، فإن المقومون يقومون بمعالجة تلك المعلومات وربطها بانطباعاتهم وفهمهم وتجاربهم وخبراتهم الخاصة لاتخاذ القرار المناسب.

بناءا على ذلك، فإن هناك احتمالية واردة أن يشوب قرار التقييم بعضا من اللاموضوعية أو ضعف في الثبات Subjectivity and Inconsistency وربما نوع من التحيز Bias. ولأن الاختبارات والمقاييس المقننة تعد أكثر موضوعية وثباتا من أدوات التقييم الأخرى، فإنها يتوقع أن تقلل من تلك الاحتمالية.

من ناحية أخرى، من المفروض أن نعرف أن التقويم لا يقتصر على الأفراد وأدائهم، بل هو نشاط يمتد إلى جوانب ومجالات متعددة كما في الكتب الدراسية، والمواد التعليمية، وطرق التدريس، والمناهج الدراسية، والبرامج التدريبية وغيرها من مدخلات العملية التعليمية كما هو مبين في المخطط المبين تاليا والذي تم اقتباسه مع بعض التعديل من نيتكو (Nitko, 2005):


يستخدم التقييم التربوي في توفير المعلومات لصالح القرارات التالية:
القرارات المتعلقة بالمناهج
القرارات المتعلقة
القرارات المتعلقة
التقويم
التقويم
السياسات
سياسات المتعلقة
تسكين
تصنيف
إدارة
اختيار
إرشاد
تخريج الطلبة
تحديد درجات
تشخيص
التغذية الراجعة
تسكين
مراقبة تقدم
تخطيط
المعلمين
الطلبة وأولياء

في ضوء ما تقدم، تبرز الحاجة إلى التمييز بين التقويم البنائي Formative والتقويم الختامي Summative (Scriven, 1967). **ويكون التقويم بنائيا (تكوينيا)**، عندما يتجه الجهد نحو:

- تقويم فعالية برنامج تدريبي أو إرشادي طور التنفيذ أو التجريب، أو تقويم كفاءة مواد تعليمية أو طرق تدريس. ويهدف هذا النوع من التقويم إلى توفير معلومات تساعد في تطوير البرنامج قبيل الانتهاء منه. كما يساعد في إجراء التعديلات أو إعادة التشكيل، أو تطوير المنتج قبيل طرحه إلى الاستخدام الميداني في الواقع. فمثلا، عند تطوير منهاج ما، فإن كلا من:
  - **الاختبارات الاستطلاعية المطبقة على عينة مناسبة من التلاميذ**
  - **قيام فريق من الخبراء لفحص محتوى المنهاج، أو**
  - **تطبيق المنهاج ميدانيا في عدد من المدارس،**

يستخدم في توفير تغذية راجعة فورية لمطوري تلك المناهج للقيام بالتعديلات اللازمة. أما دور المعلم في هذا النوع من التقويم، فيكون من خلال مراجعة الدروس أو المواد التعليمية اعتمادا على خبراته السابقة في تدريسها أو استعماله لمثل تلك المواد التعليمية من قبل.

- من ناحية أخرى، يهدف التقويم البنائي إلى تطوير أدوات القياس قبيل استخدامها في جمع البيانات من الواقع الميداني. كما يعد بعضهم الامتحانات بأنواعها والملاحظات أثناء الفصل الدراسي تقويما بنائيا.
- وعندما يكون محور الاهتمام هو الطالب، يعرف عندها التقويم البنائي بأنه عملية تشخيصية تصحيحية مستمرة، تهدف إلى ضمان إحداث التعلم المطلوب مع توفير تغذية استرجاعية لمعاودة التعليم والتقويم، وذلك بقصد تحسين مخرجات عمليتي التعلم والتعليم. كما يشير التقويم البنائي إلى نوعية تحصيله ومستوى بلوغه

أهداف معينة خلال فترة زمنية محددة. الأمر الذي أن يتوقع بموجبه أن يسهم التقويم البنائي في توجيه تعلم الطالب إلى الخبرات والمهارات المتعلقة بتحقيق هدف تعلمي لاحق.

- وعند استعراض برنامج لتقييم الحاجات Needs Assessment مثلا، يعتقد كل من إسحاق وميشيل (Isaac and Michael, 1981) أنه بعد تحديد سياسة وغايات البرنامج، يتم اشتقاق أهداف محددة قابلة للقياس وكذلك تصميم خطة تتضمن وسائل تحقيق تلك الأهداف. عند ذلك، يتم تقويم الأنشطة والاستراتيجيات، والإجراءات التي يتضمنها البرنامج موضع البحث. وهذا هو التقويم البنائي ويتألف من مكونين رئيسين هما:

1. تقويم الاستخدام Implementation Evaluation ويتم في هذا النوع من التقويم التعرف على الفجوة بين الواقع العملي للبرنامج وما يؤمل منه كما ورد في الخطة، وينتج عن هذا التقويم العمل على المحافظة على مخطط وتصميم البرنامج أو تعديله كما ينبغي.
2. تقويم الإجراءات Process Evaluation ويتم في هذا التقويم العمل على توجيه وضبط التقدم نحو الأهداف. كما يستوجب ذلك إجراء التصويبات المرحلية كما ينبغي.

ويكون التقويم ختاميا أو نهائيا أو تجميعيا، عندما يهدف إلى إصدار حكم نهائي على البرنامج ككل أو المواد التعليمية أو الإجراءات المتبعة في برنامج ما بعد نهايته. وإذا كان التقويم البنائي يستخدم في أغراض توجيهية، فإن التقويم النهائي يستخدم في إصدار الحكم النهائي على البرنامج من حيث صلاحيته لاستخدامات مستقبلية أو بقصد استبداله أو ربما إعطاء توجيهات عن كيفية التطبيق المستقبلي. وقد يخدم التقويم الختامي أهدافا مشابهة لأهداف التقويم التكويني حين يستخدم المعلم مثلا

نتائج الطلبة للعام الماضي أو لوحدة سابقة في تغذية راجعة مقبلة (Worthen and Sanders, 1987; Nitko, 1983).

وجدير بالذكر، فقد يختلف المستفيدون من نوعي التقويم. فالمعنيون في التقويم البنائي هم أولئك العاملون في تطوير البرنامج التدريبي أو المنهاج الدراسي أو في تصميم المنتج عموما (وسيلة تعليمية أو حتى سيارة أو طائرة أو جهاز كمبيوتر) وكذلك المعلمون وواضعو المناهج. بينما يشكل المستخدمون والممولون والمشرفون ودافعو الضرائب وغيرهم جمهور المعنيين بالتقويم النهائي. أو قد يشمل كذلك ما تم ذكره في التقويم التكويني.

وبالرغم من أن نوعي التقويم نشاطان متكاملان، ومتداخلان أحيانا. فإن الأمر يصبح خطرا، عندما يجري التركيز على التقويم الختامي فقط دون الاهتمام بإجراء التقويم البنائي. الأمر الذي سيجعل العملية غامضة ومحفوفة بالمخاطر، وقد يصحبها هدر في الأموال. وخير مثال على ذلك بناء منهج تربوي جديد يبنى على معايير معينة للأداء، والاسترسال في تطبيقه والعمل بدون أن تصاحب عملية التطبيق هذه عمليات تقويم لما يستلزمه من مدخلات وعمليات وما سيؤول إليه من مخرجات.

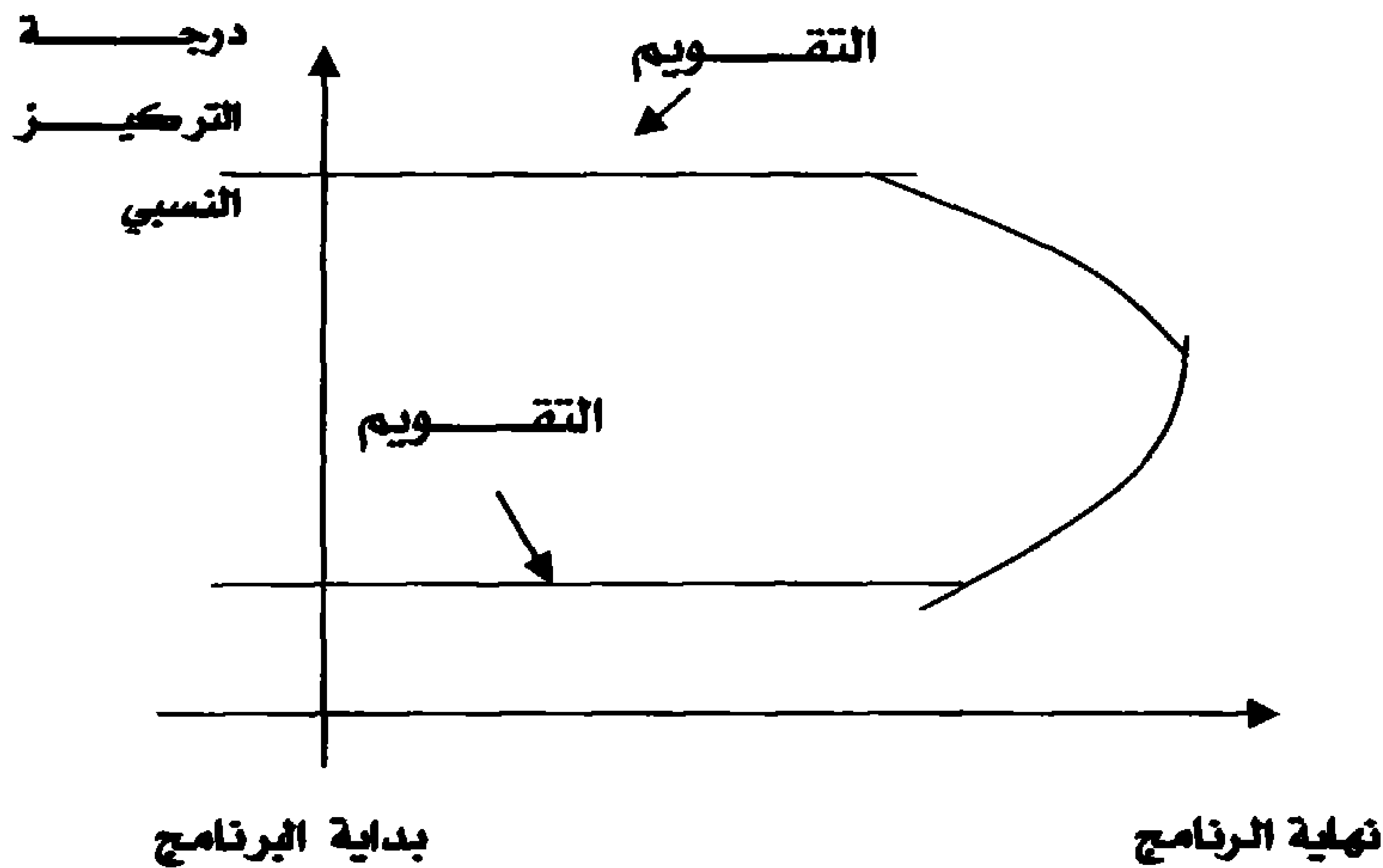
من ناحية أخرى، دعنا نتصور مدى ارتفاع كلفة تشغيل منتج ضخم كطائرة ركاب بعد الانتهاء من صناعتها دون المرور في فحص مكوناتها خطوة بعد خطوة ومرحلة بعد أخرى.

ويشير بيكر (Baker, 1978) إلى أن عاملي الوقت والتحكم أو الضبط هما الموجهان الرئيسان في استخدام التقويم البنائي بفاعلية. إذ أن المعلومات التي يتم جمعها عن حيثيات المنتج أو البرنامج موضع البحث، تكون عادة على جانب كبير من الأهمية، إذا تم إعدادها وتقديمها في الوقت المناسب. لأن تقديمها متأخرة لمتخذ القرار يجعل منها أمرا عديم الفائدة.

ويبين الجدول التالي مقارنة بين التقويم البنائي والتقويم الختامي لبرنامج معين في ضوء عدد من النقاط. وقد تم اقتباس الجدول مع إجراء بعض التعديلات من ورثن وساندرز (Worthen and Sanders, 1987):

مجال المقارنة	التقويم البنائي	التقويم الختامي
الهدف	مراقبة أو توجيه أو تحسين البرنامج	استخدام أو تطوير أو وقف العمل في البرنامج
المستهدفون	إداريو البرنامج والعاملون فيه والمستفيدون منه	الممولون ومعدو البرامج والمستخدمون
من يقوم بالمهمة	مقوم داخلي (رئيسي) مع مقوم خارجي	مقوم خارجي (رئيسي) مع مقوم داخلي
الصفة الغالبة	الجدولة الزمنية	إقناع المعنيين
أدوات القياس	رسمية وغير رسمية (ملاحظة، مقابلات، امتحانات تحريرية وشفوية وغيرها	امتحانات ومقاييس صادقة وثابتة
مرات جمع البيانات	متعددة ومتكررة حسب مراحل البرنامج	محدودة وتتم في نهاية البرنامج
حجم العينة	صغير نسبيا	كبيرا نسبيا
الأسئلة المطروحة	ما الجزء أو المرحلة التي تعمل جيدا؟ ما الذي يحتاج للتطوير؟ كيف يمكن إجراء التطوير؟ ما جدوى التغيير أو التطوير؟	ما النتائج النهائية؟ مع من وتحت أي ظروف تجري عملية التقويم ؟ بأي نوع من التدريب ؟ ما التكاليف ؟
مصدر الضعف	ما نوع البيانات المطلوب جمعها؟ ما مدى صدق وثبات أدوات القياس ؟ ما التوقيت الأمثل للتنفيذ؟	ما التساؤلات التي من الممكن إثارتها وإبرازها؟ ما محددات عمليات اتخاذ القرار ومصادرها؟

وجدير بالذكر أن يبدأ التركيز النسبي على التقويم البنائي مع بداية تجريب البرنامج موضع الاهتمام بأقصى درجة ثم يبدأ بالتناقص التدريجي، في حين يكون التركيز النسبي على التقويم الختامي في أقصى درجاته نهاية تجريب البرنامج كما في الرسم التالي الذي تم اقتباسه من ورثن وساندرز (Worthen and Sanders, 1987) بعد أن جرى إدخال بعض التعديلات:



ويعرف التقويم الختامي بأنه عملية تجميعية تقوم على تلخيص طولي عبر فترة زمنية ترمي إلى إبراز أهم نتائج التعلم المتكاملة والتي قد تضم إلى إجراءاتها إجراءات ونتائج التقويم البنائي. ويتضح مما تقدم أن نوعي التقويم نشاطان يكمل بعضهما الآخر ومستمران مع حياة البرنامج موضع البحث، ولكن بدرجات تركيز متفاوتة.

يمكن تصنيف التقويم حسب نوع المعلومات المطلوب جمعها إلى تقويم كمي وتقويم نوعي. ويعتمد التقويم الكمي Quantitative Evaluation على البيانات التي يمكن أن يعبر عنها بالأرقام كالتي نحصل

عليها من الاختبارات أو الاستبانات وبعض أدوات القياس الأخرى. وهنا يتم التعامل مع هذه البيانات من خلال مؤشرات إحصائية معروفة كمؤشرات النزعة المركزية والتشتت والمئينات.

أما التقويم النوعي Qualitative Evaluation، فيعتمد على البيانات التي نجمعها من الواقع الميداني من خلال الملاحظة وملفات ومجتمع الأتراب (الرفاق) وأولياء الأمور وغيرها. وبطبيعة الحال، يعتمد قرار التقويم عموما على نوعي البيانات ولكن ربما بدرجات تركيز تختلف حسب موضوع التقويم ونوع الأسئلة المنوي الإجابة عليها (Worthen and Sanders, 1987).

وعند الحديث عن **تقويم البرامج**، يكون الاهتمام في تقويم ثلاث مراحل في ذلك البرنامج وهي **المدخلات والعمليات والنتاجات**. ويتركز تقويم المدخلات Inputs Evaluation على عناصر البرنامج المادية والبشرية. ففي البرنامج التربوي، تعد خصائص الطلبة، والإمكانات المتوفرة في المؤسسة التربوية من تسهيلات الطباعة والتصوير ومؤهلات العاملين في البرنامج ومستوى صياغة فقرات الاختبار وغيرها.

أما **تقويم العمليات** Processes Evaluation، فهو مجموعة الإجراءات التي توجه نحو تقويم البرنامج أثناء عمليات التنفيذ، كما في طرق التدريس وطبيعة التفاعل الصفي والتوقيت الزمني لفعاليات البرنامج وخطواته، ويتضمن ذلك إجراءات سير الامتحانات وطبيعة العلاقات بين الطلبة والمعلمين وعناصر الهيئة الإدارية. ويكون تقويم النتاجات Products Evaluation عن طريق اتخاذ قرارات تتعلق بمجمل النواتج التي ستنتج كخلاصة للعمليات التي حدثت في البرنامج. ومثال ذلك مجمل أعمال الطلبة الكتابية، والمهاراتية كالرسم وكتابة التقارير وإجراء التجارب، وتسجيل المشاهدات، وتحضير الشرائح، وتطوير النماذج وغيرها.

كما قد يكون التقويم شموليا أو مكبرا Macro Evaluation أو تقويما مصغرا أو جزئيا Micro Evaluation. ويتطلب التقويم الشمولي مشاركة كافة المعنيين في البرنامج كفريق ويستغرق وقتا طويلا نسبيا ويهدف إلى تقويم البرنامج بمكوناته معا دون إغفال دور أي من هذه الجهات كالمعلمين والمديرين والمرشدين وجهات التمويل وأولياء الأمور ومصممو المناهج الدراسية ومؤلفو الكتب الدراسية وفنيو المختبرات ومنظرو الفلسفات والسياسات التربوية. بينما يهتم التقويم الجزئي بدراسة مكون واحد مثلا من تلك المكونات، كتقويم مدى فعالية استخدام المختبرات المدرسية في مجموعة مدارس ثانوية. ويختلف التقويم الشمولي عن التقويم الجزئي من حيث مقدار الجهد المطلوب بذله في عمليات كل منهما (Worthen and Sanders, 1987). وجدير بالذكر بأن الشمولية تعد نسبية، وذلك عندما يصار إلى تجزئة البرنامج الجزئي إلى أجزاء أصغر (عودة، 2000).

وقد تبرز مسألة غاية في الأهمية تتعلق بنتائج التقويم عموما وهي مشكلة عزو مسؤولية الضعف في أحد جوانب البرنامج الذي تم تقويمه. وهذا ربما يزيد الأمر تعقيدا وغموضا. الأمر الذي يجعل عملية المساءلة Accountability تسير في طريق غير يسير، خاصة عندما يغفل المقوم فحص بعض الجوانب التي لها علاقة بجوانب البرنامج موضع البحث.

وعند الحديث عن الجهة التي ستقوم بعملية التقويم، يظهر نوعان من التقويم، تقويم داخلي Internal Evaluation وتقويم خارجي External Evaluation. ويتم التقويم الداخلي من قبل مجموعة من الأشخاص من داخل البرنامج موضع التقويم، بينما يكون التقويم الخارجي نشاطا ينفذه أناس من خارج البرنامج. ولكل من هذين النشاطين سلبيات وإيجابيات.

التقييم المعياري والتقييم المحكي

تختلف طرق التقييم معياريا أو محكيا حسب طريقة تفسير النتائج على أسلوب التقييم. فهناك تفسير معياري المرجع وتفسير محكي المرجع. ففي **التفسير معياري المرجع** Norm-Referenced-Interpretation ويختصر بالرمز (**NRI**)، يتم تفسير مستوى أداء الممتحن عليها بالمقارنة مع متوسط أداء المجموعة التي ينتمي إليها وهذه المجموعة التي تعرف بالمجموعة المعيارية Norm group التي قد تكون صفا دراسيا أو مجموعة مهنية معينة أو مدرسة أو منطقة تعليمية أو على المستوى الوطني. بينما يتم في **التفسير محكية المرجع** Criterion-Reference-Interpretation ويختصر بالرمز (**CRI**) ، تفسير مستوى أداء الممتحن بالمقارنة مع مستوى أداء معين يتم تحديده مسبقا من قبل واضع الاختبار أو المؤسسة التي يتم تطبيق الاختبار لصالحها، دون النظر إلى أداء المجموعة التي ينتمي إليها الممتحن. كأن يكون المستوى الذي يحتكم إليه هو أن يحل الممتحن (85 %) من مجموع المسائل المعطاة، أو أن يقفز المتسابق مثلا (1.5) مترا ، أو أن يقوم المتسابق بطباعة (45) كلمة عربية في الدقيقة أو طباعة (55) كلمة انجليزية في الدقيقة طباعة صحيحة، على أن يكون اختيار هذه المستويات أو المعايير مبني على أساس نظري أو منفعي، أو تجريبي، أو تاريخي. ويتم خلال الفصول القادمة في هذا الكتاب مناقشة الفروق بين الاختبارات معياري المرجع ومحكية المرجع بنوع من التفصيل.

مصادر ضعف دقة القياس في العلوم السلوكية

تنبثق مشكلات القياس النفسي من خلال طبيعة الظواهر والمواقف التي يتعرض إليها القياس النفسي من ناحية. ومن خلال الخصائص التي يتصف بها القياس كما مر بداية، على أن القياس النفسي غير مباشر وغير

تام من ناحية أخرى. وتبرز المشكلة واضحة أكثر عندما يتم الحديث عن القياس في المجالات الطبيعية والمهارات الحركية، كما في القياس الفيزيائي. إذ أن القياس الفيزيائي مباشر وتام، إذ يتم تناول الظاهرة الطبيعية مباشرة لأنها موجودة فعلا وربما يسهل ملاحظتها بأكثر من أسلوب ووسيلة. لذلك يسهل تعريفها وتحديد أطرها بدقة كبيرة، وإن أي عينة صغيرة أو بسيطة ربما تفي بالغرض. هذا بعكس الظاهرة النفسية أو الاجتماعية التي قد لا يكون من السهولة ملاحظة وجودها وتلمسها، ويحتاج قياسها إلى تعريف يبين مكوناتها ودرجات تواجدها في فئة عمريه ما ضمن معطيات معينة. كما أن قياسها يحتاج إلى عناية أكبر في تكوين عينات بحثية ملائمة (Gay, Mills, and Airaison, 2009; Nunnally, 1978).

لذلك، فقد برز عدد من نقاط الضعف التي واجهت بناء وتطوير أدوات القياس النفسي عموما ، وقوائم الشخصية والميول بشكل خاص، بحيث بات من غير السهل التخلص من تلك الصعوبات. الأمر الذي دفع المعنيين إلى الاهتمام بالأدوات الإسقاطية، لتعويض القصور الذي تعاني منه أدوات القياس النفسي عموما. ولكن هذا الاتجاه نحو الأدوات الإسقاطية لم يكن بديلا كافيا وذلك لسببين رئيسين:

1. أن هذه الأدوات لم تستطع معالجة كل الظواهر النفسية.
2. أن الخصائص السيكومترية (الصدق والثبات) منخفضة نسبيا لتلك الأدوات.

ومهما يكن من أمر، فقد أمكن تلخيص أوجه القصور في دقة أدوات القياس في العلوم السلوكية في النقاط التالية:

ضعف الدافعية للإجابة: فقد تواجه مطورو الاختبارات مسألة إثارة الأفراد للإجابة على فقراتها، كما قد تتسم إجاباتهم باللامبالاة، خاصة:

- إذا لم يكن لدى المستجيبين الاهتمام الكافي، أو

- لم يتوفر الوقت الكافي، أو
- أن توقيت التطبيق لم يكن ملائما، أو
- ربما يكون الاختبار طويلا يبعث على الملل.

وبناءا على تلك الاحتمالات، فقد تكون الإجابة على الفقرات الأخيرة بسرعة وبدون أدنى انتباه أو تركيز. الأمر الذي يؤدي إلى الخروج نهاية بأداة ضعيفة لا تقوى على تقديم المساعدة المتوقعة في قياس الموقف المطلوب بصدق وفاعلية (الزوبعي وبكر والكناني، 1982).

- **التزييف في الإجابة:** ويقصد بذلك التفاوت بين السلوك اللفظي والسلوك الفعلي للفرد المستجيب. فعندما تتناول أداة القياس ظاهرة نفسية تثير الحساسية لدى المستجيب، كونها تتعلق بشخصية الفرد، فإنه يميل إلى التزييف، فقد يحاول نفي وجود الصفة أو التقليل من وجودها إذا كانت سلبية كالقلق والكآبة. أو قد يبالغ في وجود الصفة لديه إذا كانت إيجابية. وعلى الرغم من لجوء مطوري أدوات القياس إلى اتباع أساليب تكشف الزيف، إلا أنها لم تكن كافية.
- **تحكيم فعالية الفقرات:** فقد يلجأ مطورو أدوات القياس في معظم الأحيان إلى عرض فقرات مقاييسهم على هيئات من المحكمين لتقدير قوة الفقرات ومدى ملاءمتها لقياس سمة معينة. الأمر الذي ربما يؤدي إلى أن يقدم هؤلاء المحكمون بيانات متفاوتة تعكس:
  - اختلاف درجات فهم كل منهم للصفة المنوي قياسها، أو
  - يعكس رأي كل منهم المدرسة الفكرية التي ينتمي إليها، أو
  - ربما إلى اختلاف درجة اهتمام كل محكم بالأداة وبموضوعها.

كل هذا يؤدي إلى تقييم الفقرة بأكثر أو بأقل مما تستحق. الأمر الذي سيؤثر سلبا على نوعية الأداة وفعاليتها في قياس وتقييم الواقع المنشود.

- **الاختلاف في التفسير:** فقد يختلف الباحثون في تفسير النتائج التي يتم التوصل إليها جراء تطبيق أداة القياس. ويعود السبب في هذا إلى الاختلاف بينهم إلى الاختلاف في المدرسة النفسية التي ينتسب إليها كل منهم. وهذا يؤدي إلى الاختلاف في تفسير الظاهرة النفسية. الأمر الذي يجعل بعض الباحثين يركز على عوامل مسببة أو شارحة للظاهرة وإعطائها وزنا كبيرا، بينما يركز البعض الآخر على عوامل أخرى ويعطيها وزنا مختلفا.
- **صعوبة التحديد الفعلي الدقيق لكثير من المفاهيم والمكونات الوجدانية:** إذ أننا لا نستطيع تحديد الواقع الفعلي لظاهرة أو كاتجاهات فرد نحو موقف ما أو ميوله تجاه مهنة أو تخصص معين بناءا على أداة قياس واحدة. المر الذي يحتم تطوير أكثر من أداة واستخدامها لتحقيق هدف معين من ناحية، أو اللجوء إلى تطوير الأداة الواحدة وإكسابها صفة الشمول والتنوع أو أن تصبح على درجة عالية من الصدق.
- **صعوبة قياس بعض الجوانب الانفعالية**، حيث أنها ربما تتطلب ملاحظة الفرد بشكل متكرر في مواقف مختلفة يمكن أن تظهر فيها الاستجابات موضع الاهتمام. وقد تختلف الاستجابات باختلاف الموقف الذي يكون فيه الفرد. وربما تكون الاستجابات غير حقيقية أو غير طبيعية إذا شعر الفرد أنه تحت الملاحظة.
- **مناسبة أساليب الصدق المستخدمة في تطوير أدوات القياس النفسي أو التربوي وغيرها.** إذ يعد هذا البند أساسيا في الحكم على كون أداة القياس تقيس فعلا ما صممت لقياسه، فقد يمكن أحيانا إخراج أداة قياس تفتقر إلى درجة معقولة من الصدق، الأمر الذي يجعل القرارات المتخذة اعتمادا عليها ضعيفة وغير فعالة.

- **ضعف شمولية أداة القياس لمتغيرات الموقف المنوي قياسه، أو** إغفالها لبعض المجالات التي يتكون منها السمة المنوي قياسها لدى فئة من المستهدفين. وهذا يصب في صحة إجراءات صدق المحتوى المتعلقة بالاختبارات التي تمثل محالا سلوكيا معرفا كما في الاختبارات التحصيلية.
- **درجة تمثيل العينة للمجتمع المنوي تطوير أداة القياس لصالحه.** فقد لا تكون العينة التي تستخدم في تطوير أداة القياس ممثلة بدرجة كافية للمجتمع. الأمر الذي يؤدي إلى الوصول إلى تعميمات قاصرة، وتفتقر إلى المصداقية الكافية.

أسئلة وتمارين

ضع دائرة حول رمز الجواب الصحيح في الفقرات التالية:

*. أي التالية يعد مثالاً على المتغير الكمي ؟

أ- الجنس ب- نوع السيارة

ت- لون العيون ث- العمر

*. أي التالية يعد مثالاً على المتغير النوعي؟

أ- الوزن ب- معامل الذكاء

ت- الراتب الشهري ث- نوع قلم الحبر

*. ما أعلى مستوى قياس يمكن أن يصنف فيه "زمن الاسترجاع" ؟

أ-الرتبي ب- الفئوي

ت-النسبي ث- الاسمي

*. أعلى مستوى تصنف عليه أرقام لاعبي كرة القدم هو المستوى---

أ- الاسمي ب- الرتبي

ت-الفئوي ث- النسبي

*. ما أعلى مستوى يمكن أن تصنف عليه تقديرات الطلبة مثـل: (أ+، ب، ج+)

أ- الفئوي ب- النسبي

ت- الرتبي ث- الاسمي

*. إذا كانت العلاقة (س= ص ÷ ع) صحيحة، فإن أعلى مستوى يمكن أن تصنف فيه المتغيرات: س، ص، ع هو ----

أ- النسبي ب- الفئوي

ت - الرتبي ث- الإسمي

أجب بـ "نعم" أو "لا" على العبارات التالية:

*. أقصى مستوى قياس يمكن أن يقاس عليه متغير الطول هو الفئوي.

*. أقصى مستوى يمكن أن يقاس عليه متغير التحصيل الدراسي هو الرتبي.

*. تعتبر جنسية الفرد متغيرا نسبيا.

*. يعتبر القياس الصفي قياسا تربويا فقط.

*. يعتبر القياس نشاطا كميا بينما التقويم نشاطا نوعيا.

*. التقييم أعم من القياس وأشمل من التقويم.

*. الاختبار غير الرسمي هو اختبار مقنن يستخدم في تقييم مستوى تحصيل الطلبة.

*. التقويم التجميعي نشاط يهدف إلى الحكم على نجاح البرنامج خلال تجريبه.

*. القياس في العلوم السلوكية غير تام لوجود أخطاء شخصية عند جمع البيانات.

*. يكون حجم العينة في حالة التقويم التجميعي أكبر منه في حالة التقويم التكويني.

*. القياس في العلوم السلوكية نسبي لعدم وجود صفر حقيقي في متغيراتها.

*. يعد تفسير الدرجات على اختبار محكي المرجع الأفضل عند ترتيب المتنافسين.

*. إذا كان جميع أفراد عينة دراسة يحملون الدكتوراة إن المؤهل العلمي يعد ثابتا.

*. الرقم الجامعي هو متغير نوعي وإسمي وتصنيفي وكمي.

*. إن اختلاف الأفراد في مستوى دخولهم يشير إلى أن الدخل متغير يقع علىمستوى القياس الاسمي.

*. إذا كان دخل أستاذ الجامعة أفضل من دخل المهندس، فهذا يعني أن الدخل متغير يمكن أن يقع في المستوى الفئوي.

*. إذا كان (ص=3 س)، فإن أعلى مستوى يمكن أن يقع عليه (س)، (ص) هو الفئوي.

*. إن قياس بعض السمات يعد مضيعة للوقت بسبب ضعف ثباتها.

*. إذا كان (أ) أكبر من (ب)، فإن أعلى مستوى يمكن أن يقع عليه (أ) و (ب) هو المستوى الاسمي.

املأ الفراغات التالية:

*. في مستوى القياس ـــــــ فإن مجموع المقدارين (100) و (50) يساوي (150).

*. إن الإكتفاء بأربعين فقرة لقياس القدرة العددية يشير إلى أن القيـــاس في العلوم السلوكية ـــــ.

*. إن استخدام الأداء على استبانة لقياس مستوى الإكتئاب لدى مجموعة من الشباب يشير إلى أن القياس في العلوم السلوكية ـــــ.

الفصل الثاني

بناء وتطوير أدوات القياس

Constructing and Developing the Measurement Tools

- مقدمة
- خطوات بناء أداة القياس
  - تحديد أهداف استخدام الدرجات على أداة القياس
  - تحديد عناصر المحتوى أو جوانب السلوك التي تمثل السمة
  - بناء جدول المواصفات
  - بناء فقرات المقياس
  - مراجعة فقرات المقياس
- قواعد عامة لكتابة فقرات الإختبار
  - الفقرات الانتقائية
  - الفقرات الصياغية
- بناء الإختبارات محكية المرجع
- نظرة إضافية نحو تخطيط الإختبار
- ملخص
- أسئلة وتمارين

2

المخرجات التعليمية

بعد دراسة هذا الفصل يتوقع من الطالب أن يكون قادرا على أن

- يصف أهمية بناء أداة القياس.
- يصف الظروف والشروط المؤثرة بأداة القياس.
- يعدد خطوات بناء أداة القياس.
- يصف مجالا سلوكيا لمقياس في مجال قدرة عقلية.
- يصف الأنشطة التي تستخدم في تحديد السلوكات المرتبطة بالسمة.
- يبني جدولا لمواصفات اختبار تحصيلي.
- يعدد أشكال فقرات الإختبارات.
- يصف خصائص المموه الجيد في الفقرة الموضوعية.
- يبني فقرة موضوعية من نوع الاختيار من متعدد مراعيا لشروط الصياغة.
- يبني فقرة موضوعية من نوع المزاوجة مراعيا لشروط الصياغة.
- يبني فقرة من نوع التكميل مراعيا شروط صياغة فقرة جيدة.
- يبني فقرة مقالية مراعيا شروط صياغة فقرة جيدة.
- يحسب درجة مفحوص بعد تصحيحها لأثر التخمين.
- يعرف مفهوم حكمة الإختبار.
- يصف حالة قلق الإختبار.

مقدمة

مضى على حركة بناء المقاييس والإختبارات زمنا طويلا، ليس ذلك فيما يتعلق بعدد هذه الأدوات وتنوعها فحسب، بل من حيث اعتماد الطرق المنهجية في تطوير عباراتها أو فقراتها من ناحية، وتسهيل عمليات التطبيق والتصحيح والتفسير من ناحية أخرى. ومهما يكن من أمر، فإن لكل أداة خصوصية في التصميم والتطبيق وطريقة التفسير. وهذا يحتم ضرورة الإنتباه لمجموعة المبادىء والشروط التالية قبيل الشروع ببناء أي أداة، ومن ثم يصار إلى دراستها بعناية فائقة من أجل تعميق مستوى الإتقان في الخروج بأداة أكثر فاعلية، ومصداقية، ودقة، وعملية، وموضوعية:

استطلاع الأهداف المتوقع تحقيقها جراء تطبيق أداة القياس

تختلف أدوات القياس باختلاف الهدف الذي تسعى إلى تحقيقه. فبعض الأهداف يناسبها الإختبار التحريري الجماعي كأدوات التقدير وقوائم الفحص، في حين هناك أدوات تهتم بتقويم مستوى الأداء في المهارات كما في اختبار رسم الرجل وغيره.

من ناحية أخرى، يستلزم تحقيق بعض الأهداف استخدام أدوات إسقاطية، وبعضها يتطلب استخدام أدوات محددة البناء (يكون المثير والإستجابة فيها محددين). وهناك أدوات تهدف إلى تقييم حالة قائمة من قدرات الفرد وأخرى تهتم بتقييم درجة استعداد الفرد نحو ممارسة مهارة أو الاشتغال بوظيفة معينة مستقبلا.

استطلاع طبيعة المجال السلوكي موضع الاهتمام وتحديد مكوناته

عندما تقيس الأداة مجالا سلوكيا محددا ومعرفا (ضرب الكسور العادية وقسمتها)، له مرجعية معروفة كمنهاج دراسي أو كتاب (كتاب

الرياضيات لمستوى الصف الخامس الابتدائي) تكون تلك الأداة اختبارا تحصيليا، بينما تكون تلك الأداة اختبار قدرة أو اختبار استعداد عندما لا يكون المجال محددا بأطر واضحة ومعروفة. وخير مثال على ذلك اختبارات الذكاء والاستعداد والابتكار. أما عندما تكون الأداة اختبارا أو مقياسا نفسيا (كمقياس القلق، أو قائمة شخصية، أو نحو ذلك)، فليس من السهولة بناؤها لعدم توفر مرجعية سلوكية محددة كما في المنهاج الدراسي، أو المادة التعليمية المعرفة.

استيعاب خصائص المستهدفين بأداة القياس

وهذا يتعلق بكون الأداة المطلوب بناؤها أو تطبيقها على المفحوصين كل على حدة (أدوات فردية) أو تطبيقها على مجموعة منهم معا في صف أو مدرسة أو في منطقة تعليمية (أدوات جمعية). كما يتعلق ذلك بالفئة أو الفئات العمرية للمستهدفين. إذ يترتب على ذلك تحديد الزمن الذي يخصص لتطبيق أداة القياس من ناحية، ونوع فقرات الأداة من ناحية أخرى، وتحديد شكل الأداء المطلوب. فأطفال المرحلة الابتدائية الأولى أو رياض الأطفال لا يمكنهم الجلوس لاختبار ما يستغرق أكثر من 15-20 دقيقة، كما أنه لا يمكنهم الجلوس للاستجابة على فقرات يحتاج الإجابة عليها تعليمات مفصلة أو مطولة، كالفقرات الانتقائية. إضافة إلى أن فقرات المقاييس المخصصة للأطفال الصغار غالبا ما تكون مصوغة على هيئة صور وأشكال (أدوات غير لفظية)، في حين تبنى أدوات القياس اللفظية التي يحتاج الإجابة عليها ورقة وقلم لصالح الممتحنين القادرين على القراءة والكتابة.

مدى توفر الإمكانات والتسهيلات المتعلقة بتطبيق وتصحيح أداة القياس

فقد تتوفر وسائل متنوعة (يدوية، آلية، أو حاسوبية) تستخدم في إخراج وتطبيق وتصحيح الإجابة على أدوات القياس، الأمر الذي ربما يشجع بناء أدوات قياس من نوع الاختيار من متعدد من ناحية، وتطبيق تلك الأدوات على عدد غير قليل من الممتحنين من ناحية أخرى. في الوقت الذي ربما يؤدي عدم وجود مثل تلك الإمكانات إلى التفكير ببناء فقرات من نوع الإجابة المصوغة من زاوية، وإلى الاعتماد على عينات محدودة من الممتحنين من زاوية أخرى.

من هنا، تبرز أهمية المعلومات **Knowledge** والمهارات **Skills** اللازمة لتطوير أدوات القياس عموما والإختبارات المتعلقة بالجوانب المعرفية والتحصيلية على وجه الخصوص على النحو المطلوب لأنها:

(1) فرصة جيدة للدارسين في التعرف على أنواع السلوك والجوانب المعرفية التي يجب التركيز عليها في دراستهم. إذ أن هذا يعد بمثابة تعريف إجرائي محدد للأهداف التعليمية التي يهدف المنهاج إلى تحقيقها.

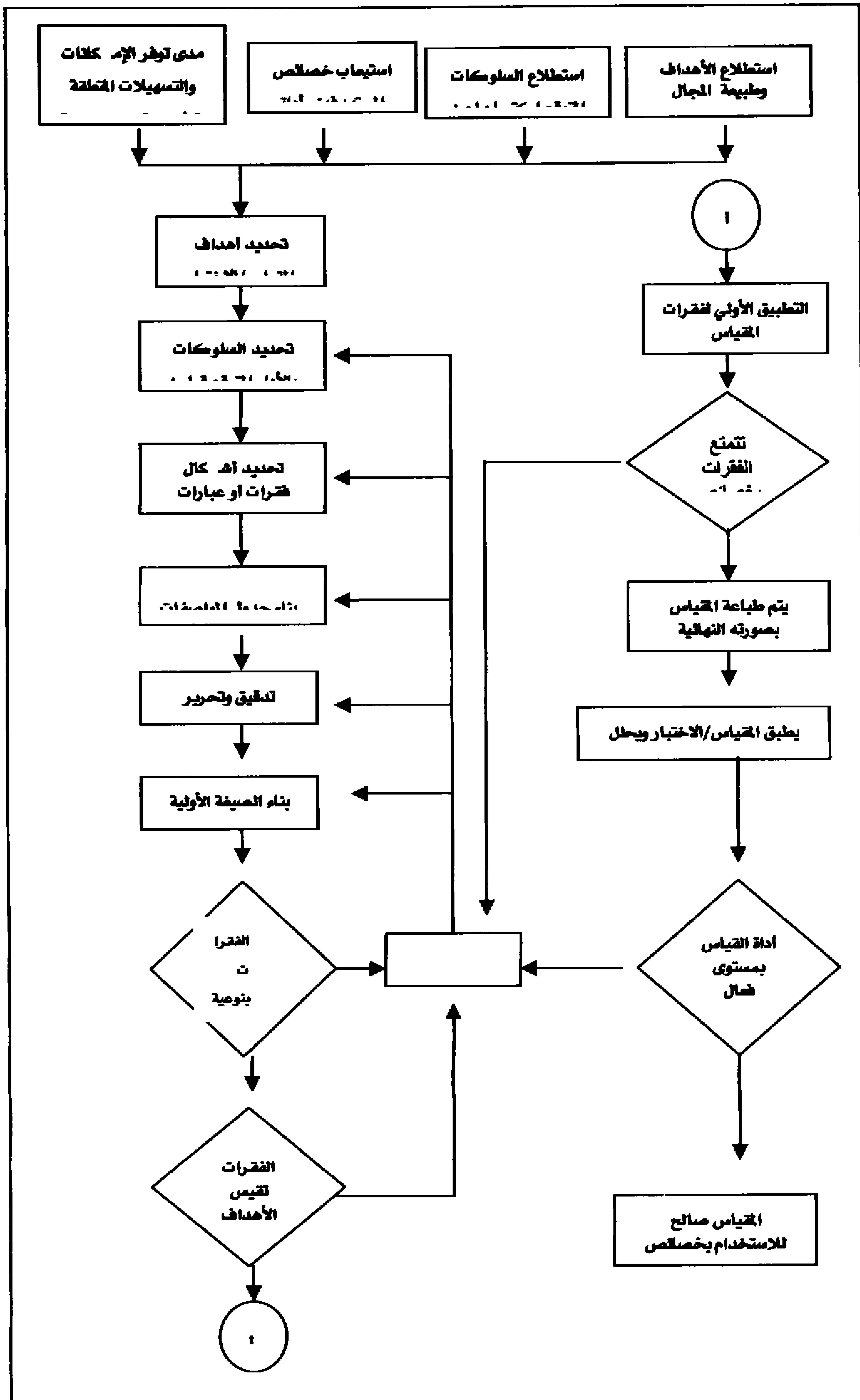
(2) وسيلة تساعد في النهوض بعمليات تصميم وتخطيط المناهج، ومن ثم تطويرها. كيف لا، والتعلم هو عملية إحداث تغيير مقصود في السلوك، ويبدأ الإختبار بتحديد مدى حدوث هذا التغير وضمان نوعيته.

(3) تدفع باتجاه تقويم وتحسين ما هو متوفر من أدوات قياس في الأسواق، فقد نجد عددا ليس قليلا من الأدوات المتوافرة في متناول المعلمين أو الطلبة بقصد المساعدة في تحسين أدائهم.

(4) تؤدي إلى إجراءات أفضل موضوعية وأكثر عدالة **fair** في قياس وتقويم أداء الطلبة، إذ أن التلاميذ هم الجهة الأكثر استهدافا بعملية

التقويم التي يقودها المعلمون. كما أن الموضوعية والعدالة لا تتحققا بناءا على أدوات قياس يتم تطويرها من قبل أناس غير محترفين، أو الذين لا يمتلكون معلومات ومهارات لازمة **(Nitko, 1983)**.

وبالختام، يمكن مطالعة النموذج التالي الذي يبين طريقة مقترحة لفهم كيفية بناء المقاييس عموما، والإختبارات التحصيلية علـــى وجـــه الخصوص كما هو في الشكل التالي:


مدى توفر الإم كانات
والتسهيلات المتعلقة
استيعاب خصائص
استطلاع السلوكات
استطلاع الأهداف
وطبيعة المجال
تحديد اهداف
تحديد السلوكات
تحديد أشـ كال
فقرات او عبارات
تدقيق وتحرير
بناء الصيغة الأولية
الفقرا
ت
بنوعية
الفقرات
تقيس
الأهداف
التطبيق الأولي لفقرات
المقياس
تتمتع
الفقرات
يتم طباعة المقياس
بصورته النهائية
يطبق المقياس/الاختبار ويحلل
أداة القياس
بمستوى
فعال
المقياس صالح
للاستخدام بخصائص

مراحل بناء أداة القياس

على الرغم من ظاهرة الاتساع في انتشار أدوات القياس كما وتنوعا، فإن هناك خطوات رئيسية يمكن تتبع معظمها لبناء أدوات القياس وهي على النحو التالي موضحا إمكانية الرجوع إلى المخطط الاسترشادي لخطوات بناء وتطوير أداة القياس كما ظهرت في الشكل السابق.

1- تحديد أهداف استخدام الدرجات على أداة القياس

تعتبر عملية تحديد أهداف المقياس المنوي بناؤه خطوة أساسية في بنائه حسب الأصول، مع ضرورة تحديد الأهمية النسبية لكل هدف من هذه الأهداف. ولا يخفى أن عملية اشتقاق الأهداف وصياغتها تتفاوت من حيث الوضوح واليسر حسب نوع الظاهرة أو السمة التي من المتوقع أن يتناولها المقياس. إذ أن صياغة الأهداف بالنسبة للاختبارات التحصيلية أكثر سهولة من تلك المتعلقة بالمقاييس النفسية، لأنها تتميز بوجود منهاج دراسي محدد يحوي مادة تعليمية لها أبعاد محددة وواضحة المعالم، فيتم كتابة وصياغة أهداف المقياس وفق أهداف تدريس المادة التعليمية. بينما يتطلب الأمر غير ذلك عند الحديث عن المقاييس النفسية، إذ يتطلب ذلك تعريف محدد للظاهرة النفسية التي سيتعرض المقياس لها.

يكون هذا التعريف اصطلاحيا ، وهو التعريف الذي يتم الحصول عليه بالرجوع إلى الموسوعات العلمية المتخصصة أو القواميس أو النظريات والكتب ذات العلاقة، أو تعريفا إجرائيا وهو التعريف الذي يتم اقتراحه وتبنيه كما يتطلب واقع الموضوع المنوي قياسه، ويقوم بهذا مصمم المقياس بالتعاون مع فريق الخبراء والمتعاونين. وفي كل الأحوال يفترض أن تصاغ أهداف المقياس بصورة تفصيلية قابلة للقياس والملاحظة وترتبط ارتباطا وثيقا بالنشاط المتوقع تحققه.

ومن المفروض أن تتم عملية بناء وتطوير المقاييس في ضوء الأهداف الأساسية التي تشتق من طريقة تفسير واستخدام الدرجات على الإختبارات أو المقاييس. فلو اتجهت النية لتطوير اختبار لقياس الإستيعاب القرائي **reading comprehension** مثلا لدى الطلبة المتوقع التحاقهم في الجامعة، فإن المعلومات التي سيتم جمعها عن هذا الإختبار قد تستخدم في قرارات قبول **admission**، أو تعيين **placement**، أو تشخيص **diagnosis**. ومن غير المعقول أن تستخدم الدرجات على الإختبار الواحد في عدة عمليات مجتمعة. لأن الإختبار الذي سيستخدم لأغراض التمييز بين الطلبة في امتلاكهم لقدرة ما تكون فقراته متوسطة الصعوبة. من ناحية أخرى، يستخدم الإختبار التشخيصي في استقصاء مواطن ضعف وجوانب قوة الأفراد ممن هم على مستوى قدرة منخفض. وعليه يتوقع أن تكون فقرات تلك الإختبارات سهلة نسبيا.

وبشكل محدد، فإن الإختبار الذي تستخدم نتائجه في عمليات تعيين أو تصنيف مجموعة من المتنافسين المحترفين يختلف حتما في محتواه وخصائصه عن ذلك الذي يستخدم في اكتشاف خصائص الأشخاص الذين هم بحاجة إلى برامج تأهيل خاصة أو إعادة تدريب.

2- تحديد عناصر المحتوى أو جوانب السلوك التي تمثل السمة

تبرز أهمية هذه الخطوة في كونها الأساس الذي تبنى عليه الفقرات، والمجال الذي تشتق منه. فبعد تعريف الظاهرة (السمة)، يتم تحديد أهداف المقياس، و يصار إلى تحديد مكونات تلك الظاهرة، أو يلجأ إلى تقسيم الظاهرة إلى عناصرها الأولية، حيث يمثل كل عنصر مجالا معينا، أو إطارا مرجعيا، يتم اشتقاق الفقرات وكتابتها منه. ومهما يكن من أمر، فما زالت عملية ترجمة السمة النفسية إلى عدد من فقرات الإختبار أو العبارات تعد عملا خاصا **private** ، وغير رسمي **informal** ، وغير موثق

undocumented (Crocker and Algina, 1986). وهناك عدد غير قليل من المهتمين في ذلك من أمثـــال كرونبـــاخ **(Cronbach, 1970)** وهالادايـنـــا **(Haladyna, 1980)** وشوميكر **(Shoemaker, 1975)** الذين اشتغلوا في هـــذه المسألة رغم تركيزهم في الأساس على الإختبارات التحـــصيلية. والـــشائع حاليا هو أن يبدأ مصمم أداة القياس بافتراض عدد من أنواع السلوك الدالة على السمة المنوي قياسها، والتي يعتقد أنها تمثل تلك السمة، ليصار إلـــى بناء عدد من الفقرات لهذا الغرض. وقد يقود هذا الإجراء إلى حذف عـــدد من المجالات المهمة أو التركيز على مجالات لا تتعلـــق بالـــسمة بدرجـــة كبيرة، ذلك لأن هامشا كبيرا من ذاتية أو شخـــصية أو ثقافـــة أو خلفيـــة الشخص القائم على بناء الإختبار ربما يؤثر في تلك العملية. ولضبط العمل والتحكم به على أفضل وجه، لا بد لمطور أداة القياس من الاهتمـــام فـــي واحد أو أكثر من الأنشطة التالية:

1.2: تحليل الاستجابات على أسئلة أو مثيرات مفتوحة النهاية Responses Analysis فقد يمكن لمطور أداة قياس سمة معينة أن يبدأ عمله بطرح أسئلة مفتوحة على الفئة المستهدفة، ويقوم بعدها بتصنيف الاستجابات وتحليلهـــا، ومن ثم يشتق الأبعاد الرئيسة المكونة للسمة المنوي قياسها. وخيـــر مثـــال على ذلك ما تم تطبيقه على مجموعة من الأطفال الذين تم تحليل كتابـــاتهم في مواضع التعبير للوصول إلى مقياس لمفهوم الذات لديهم عـــن طريـــق اشتقاق الأبعاد التي تكونت منها الأداة.

من ناحية أخرى وفي حالة الإختبار التحصيلي، يتم اشتقاق وتطـــوير مموهات الفقرات الانتقائية، أو تطوير قواعد تقدير **Rubrics**، للحكم علـــى مدى قدرة الطلبة على التعبير، ليساعد كل ذلك في تطوير أداة تقييم مناسبة.

2.2: مراجعة البحوث والدراسات المتعلقة بموضــوع الــسمة Review of Research وهذا ربما يساعد في تكوين فهم أوسع لمكونات السمة وعلاقتها بسمات أخرى، إضافة إلى المساعدة في تعريف شامل مفــصل لهــا مــن ناحية، أو لاقتراح مخطــط منهجــي يــساعد فــي تطــوير الخــصائص السيكومترية للأداة ككل، ولمكوناتها من ناحية أخرى.

3.2: المواقف الطرفية Critical Incidents كأن يسأل مصمم أداة القيــاس مجموعة المشرفين على مهنة معينة أو وظيفة ما عن وصف المواقف التي تكون فيها فعالية العامل أقصى ما يمكن، أو تلك التي يكون فيها الأداء بأقل مستوى، وهذا يساعد في تكوين المتصل الكمي والمنطقي للمواقــف التــي تكون السمة، ليسهل قياسها وتحديد عناصرها (Flanagan, 1954). وربمــا تنطبق هذه الخطوة على بناء المقاييس (النفسية أو الاجتماعية أو الشخصية) أكثر منها على بناء الإختبارات (التحصيلية أو العقلية) عموما.

إذ يعمل تبعا لذلك مصمم أداة القياس لتطوير عبارات تمثل وصف الــسمة لدى أفراد فئة عمرية معينة بحيث تكون بمجملهــا متــصلا **Continuum** منطقيا لحدي تلك السمة.

4.2: الملاحظات المباشرة Direct Observations فقد يتعرف مطور أداة القياس على أنواع السلوك المكونة لسمة القلق مثلا، من خــلال ملاحظــة المرشد **Counselor** لمستوى القلق المرتبط بمهنة أو وظيفة كما يبدو علــى ملامح المنخرطين فيها.

5.2: استشارة الخبراء Experts Judgment إذ يمكن لمصمم أداة القياس الحصول على مدخلات **Inputs** مهمة من الخبراء المشتغلين بالــسمة، فقــد يمكن جمع بيانات حول السمة، كمفهومها ومكوناتها وغير ذلك من خــلال المقابلات والاستبانات وقوائم الشطب أو سلالم التقدير. وخير مثــال علــى

ذلك عندما يتم تجميع الأعمال التي يتوقع من الممرض مثلا أن يقوم بها في المستشفى عن طريق توزيع عدد من الأسئلة على المعنيين في مستشفى كبير لمسح المهام والوظائف التي يقوم بها الممرض مثلا في أوقات متعددة ليصار في النهاية إلى تصنيف هذه المهام وترتيبها وفق سياق معين، والتي ستشكل في شكلها الأخير أداة تقييم عمل الممرض.

6.2: الأهداف التدريسية Instructional Objectives وهنا يسأل المتخصصون في تدريس المواد الدراسية أو من سيقوم بتقويم مستوى أداء الطلبة فيها لمراجعة موادهم الدراسية، ومن ثم يتم تطوير مجموعة من الأهداف التدريسية للاختبار (في حالة الإختبارات التحصيلية). وتعرف هذه الصيغ بالأهداف السلوكية التي يتوقع منها أن تصف سلوكا قابلا للملاحظة والقياس في الأوضاع الصفية، مع مراعاة نوع المحتوى، ومبتعدا عن الصياغة التي لا تمكن مصمم الإختبار من تقويم الأداء في الصف الدراسي.

وبكل الأحوال، فإنه يجب تحديد الأهمية النسبية لكل عنصر أو بعد من أبعاد الظاهرة، أو السمة النفسية، أو القدرة العقلية، أو حتى التحصيل في المادة الدراسية. ففي حالة اختبار لقياس المستوى التحصيلي لطلبة صف ما في مادة اللغة العربية مثلا، فإن ذلك يستدعي تقسيم المحتوى إلى الأجزاء التي يتكون منها (كالأسماء الخمسة، والأفعال الخمسة، وكان وأخواتها، والمبتدأ والخبر، وحروف الجر). وأن يتم بعد ذلك تحديد الأهمية النسبية لكل جزء مثل (30%، 20%، 15%، 25%، 10% على الترتيب). ويتم تحديد تلك الأهمية في ضوء عدد من العوامل مثل:

- الوقت الذي استغرق في تدريس هذا الجزء من المادة التعليمية،
- أهمية كل جزء من أجزاء المحتوى لأغراض التعلم اللاحق،
- حجم المادة التعليمية في المنهاج التعليمي،

- مقدار ما يشرحه الجزء من مجمل التباين في السمة، وهذا يأتي نتيجة لبعض المعالجات الإحصائية مثل التحليل العاملي.

3- بناء جدول المواصفات

يمثل جدول المواصفات **Table of Specifications** مخططا لتوزيع فقرات أداة القياس في ضوء توزيع أقسام المحتوى السلوكي ومجموعة الأهداف التي يقيسها الإختبار. وهنا تظهر أهمية كل جزء من أجزاء المدى السلوكي للسمة وكل هدف يتوقع تحقيقه. ولاشك أن الأهداف تصنف في بعد معرفي، أو وجداني، أو مهاري). ومن المفروض أن يكون هناك توازنا في توزيع الفقرات على الأهداف في المستوى المعرفي من جهة، وعلى أقسام المحتوى (المدى السلوكي) من جهة أخرى.

وتبرز الفائدة الرئيسة المتوخاة من بناء جدول المواصفات في ضرورة أن تمثل أداة القياس المحتوى الدراسي (في حالة اختبار التحصيل) أو المدى السلوكي (في حالة الإختبارات القدرات أو السمات) بأفضل صورة ممكنة. كما يمكن أن يساعد جدول المواصفات في إعطاء صورة مقنعة ومفصلة لأداة القياس تفيد مراجعي الإختبار و مستخدميه (**Crocker and Algina, 1986**). وفي حالة الإختبارات التحصيلية، تقع الأهداف في المستوى المعرفي في ستة مستويات حسب ما جاء في تصنيف بلوم **(Bloom, 1965)** هي المعرفة، والاستيعاب، والتطبيق، والتحليل، والتركيب، والتقويم:

- **المعرفة knowledge**: ويتلخص في القدرة على الاسترجاع أو التذكر أو التعرف على الحقائق والمفردات المتعلمة بنفس الطريقة التي سبق عرضها وتعلمها. ومثال ذلك استدعاء عواصم بعض الأقطار ومؤلفو عدد من الكتب أو الروايات أو رموز العناصر والمركبات الكيماوية.

مثال: أن يكون الطلب قادرا على كتابة تعريف للقياس بناءا على ما تم قراءته في كتاب أساسيات القياس في العلوم السلوكية.

- **الاستيعاب Comprehension**: وهو أدنى مستوى للفهم، ومن أكثر مستويات القدرة الذكائية شيوعا (عودة، 2000). وتكون على هيئة: ترجمة، وتفسير، واستكمال:
- **ترجمة Translation** المفاهيم والأفكار أو التعبير عنها بشكل يختلف عن ما هو عرض أصلا باستخدام كلمات جديدة أو رموز أو صورا أو أشكالا.
- **تفسير Interpretation** وهي عملية إدراك العلاقات الواردة في مقدمة ما كتفسير البيانات الواردة في شكل أو رسم بياني معين.
- **استكمال Extrapolation** ويتمثل في تقديم استنتاجات وتنبؤات بعد استقراء معلومات معطاة في موقف معين.
- **التطبيق Application**، وهي عملية استخدام المعلومات والبيانات السابقة لحل مشكلة معينة. والاختلاف بين التطبيق والاستيعاب يتلخص في أن مستوى التطبيق يتميز بعدم وجود أية إشارة لتلك المعلومات في مقدمة السؤال. وجدير بالذكر أن الشخص لا يستطيع توظيف معلومات سبق له معرفتها بدون أن يكون قد استوعبها وفهمها جيدا مسبقا.

مثال: أن يكون الطالب قادرا على حل مسألة تحتاج مبدأ حل المعادلة بمجهولين.

- **لتحليل Analysis**، هو القدرة على التعرف على وجود علاقة أو أكثر بين عدد من الحقائق والفرضيات والاستنتاجات الواردة في موقف معين، أو تجربة ما، أو قطعة أدبية، أو مجموعة من البيانات والمعطيات. وما يميز هذا المستوى أن المعاني المشتقة غير صريحة

وغير مباشرة. والمطلوب في التحليل أكثر مما هو كذلك في الاستيعاب.

مثال: أن يكون الطالب قادرا على تصنيف المواد مغناطيسيا حسب درجة انجذابها نحو المغناطيس تحت درجات حرارة متنوعة.

- **التركيب Synthesis**، هو قدرة الشخص على القيام بإنتاج فريد ومميز ككتابة موضوع حول قضية ما أو رسم خطة أو اشتقاق علاقات مجردة غير واردة في النص. أي أن يتم تقديم عملا إبداعيا.

مثال (1): أن يكون الطالب في الصف العاشر قادرا على تصميم تجربة تبين علاقة حجم الغاز بضغطه عند ثبات درجة حرارته باستخدام خامات بيئية متاحة.

مثال (2): أن يكون الطالب قادرا على تصميم وتخطيط تمثيلية تهدف إلى ضرورة الحفاظ على البيئة أو تنمي الوعي البيئي لدى الأطفال.

- **التقويم Evaluation**، وهو القدرة على اتخاذ حكم أو تثمين قرار تجاه مسألة اجتماعية (قضايا المرأة والانتخابات، أو قضايا الإنجاب والتنظيم الأسري)، أو مسألة تلوث البيئة والوعي البيئي والوقود الحيوي، أو قضية الحفاظ على الطاقة وترشيد استهلاكها. وتكمن أهمية مستوى التقويم في كونه حلقة وصل بين المجال المعرفي والمجال الانفعالي لكونه نشاطا يستوجب إصدار حكم أو قيمة والذي يعد أحد مستويات المجال الانفعالي.

والجدول التالي يمثل جدول مواصفات لاختبار يغطي وحدة دراسية في مادة العلوم العامة موضوعها القوى (Nitko, 1983). وقد مثلت الصفوف أقسام محتوى الوحدة الدراسية، بينما مثلت أعمدة الجدول الأهداف التي يتوقع أن تتحقق لدى المتعلم حسب تصنيفات بلوم للأهداف السلوكية:

جدول مواصفات لاختبار في العلوم في وحدة القوى

	المعرفة	الاستيعاب	التطبيق	التركيب	المجموع
المفهوم التاريخي للقوى	%100 (2)				% 5 (2)
أنواع القوى	% 50 (2)	% 50 (2)			% 20 (4)
قوى تعمل ببعدين	% 25 (2)		% 75 (6)		%25 (8)
قوى تعمل بثلاثة أبعاد	% 10 (1)		% 90 (9)		%25 (10)
تفاعل الكتل	% 40 (6)		% 50 (8)	% 10 (2)	% 40 (16)
المجموع	13	2	23	2	%100 (40)

ويتضح من هذا الجدول أن الإختبار موضع البحث والتطوير يتألف من 40 فقرة تغطي خمسة أقسام من وحدة القوى، كل قسم يتمتع بأهمية نسبية مناسبة. إذ يبدو أن الجزء "تفاعل الكتل" يحظى بالأهمية القصوى (40 %) من بين أقسام الوحدة الدراسية. وأن هناك 16 فقرة يجب بناؤها من هذا الجزء، في الوقت الذي لا يحظى قسم المفهوم التاريخي للقوى بأكثر من 5 % من مجموع الفقرات بواقع فقرتين فقط. كما يلاحظ أنه ليس من الضرورة أن تتوزع الفقرات على كافة مستويات بلوم، فقد تم توزيع فقرات تفاعل الكتل على ثلاث مستويات هي المعرفة، التطبيق، والتركيب. في حين توزعت معظم أجزاء الوحدة الباقية على مستويين فقط. ولا يقتصر جدول المواصفات على الإختبار التحصيلي، بل يمكن تعميمه على المقاييس النفسية أو الإختبارات العقلية وحتى قوائم الشخصية. إذ يتوقع من

تكوين جدول مواصفات متوازن إلى الخروج بأداة تتمتع بدرجة عالية من صدق المحتوى. ويبين الجدول التالي نموذجا لجدول مواصفات لمقياس يستخدم في تقييم مستوى تأهيل المعلمين في ولاية فلوريدا الأمريكية (Crocker and Algina, 1986).

جدول مواصفات لمقياس لتقييم برنامج تأهيل المعلمين في ولاية فلوريدا الأمريكية

	مستوى الهدف المعرفي			
أقسام محتوى الإختبار وأرقام كفايات كل منها	المعرفة	التطبيق	حل المسألة	المجموع
الإدارة الصفية: الكفايات 6، 12، 13، 15، 16، 17، 20، 22	1	10	9	20
تنمية الطلبة والارتقاء بهم: الكفايات6، 11، 10، 12، 16، 17، 20، 21، 22، 23	1	11	11	23
تقويم وتوثيق تقدم الطلبة: الكفايات 6، 7، 10، 11، 12، 14، 16	2	16	1	19
المواد التعليمية: الكفايات 10، 12، 13، 15، 21	1	6	2	9
الأهداف التدريسية: الكفايات 9، 10، 11، 12، 14	---	7	3	10
التعلم والتعليم: الكفايات 6، 9، 10، 11، 12، 13، 15، 17، 21، 22، 23	2	12	5	19
المجموع	7	62	31	100

يلاحظ أن المقياس يغطي ستة مجالات تتفاوت في أهمية كل منها، ويتكون المقياس من 100 فقرة من نوع الاختيار من متعدد، توزعت على ثلاثة مستويات معرفية هي المعرفة والتطبيق وحل المسألة. وهذه بالضبط ليست كلها من الأهداف المعرفية الواردة في تصنيف بلوم. كما يلاحظ أن غالبية الفقرات(62 %) تصنف في مستوى التطبيق. وأن تركيز الفقرات (16) كان في مستوى التطبيق وتنتمي إلى المجال " تقويم وتوثيق تقدم الطلبة".

والجدير بالذكر فقد تمت الإشارة هنا إلى أرقام الكفايات **Competencies** التي تنتمي إليها أقسام محتوى ذلك المقياس، وهذا يختلف عن ما ورد في الجدول السابق عندما كان الحديث مقتصرا على مواصفات الإختبارات التحصيلية.

أما الجدول التالي، فهو جدول لمواصفات اختبار نهاية الفصل يقيس مستوى تحصيل طلبة الصف الثامن في مادة الصحة (**Thorndike and Hagen,1986**). تكون الجدول من بعدين، **أولهما** يبين مجالات المحتوى الثلاثة (التغذية ، الأمراض المعدية، الأمراض غير المعدية)، **وثانيهما** يستعرض الأهداف كعمليات ، الثلاثة الأولى تتدرج تحت مستوى المعرفة، ويصنف الهدف الرابع في مستوى التقويم ، بينما يصنف الخامس على أنه تطبيق.

جدول مواصفات لامتحان نهائي في مادة الصحة للصف الثامن

الأهداف كعمليات		مجالات المحتوى			مجموع الفقرات
		التغذية 40 %	الأمراض المعدية 40 %	الأمراض غير المعدية 20 %	
معرفة	يتعرف على مصطلحات ومفردات (20%)	5= 4.8	5= 4.8	2=2.4	12
	يتعرف على حقائق نوعية (30 %)	7= 7.2	7= 7.2	4 = 3.6	18
	يتعرف على المبادىء والتعميمات (30%)	7= 7.2	7= 7.2	4 = 3.6	18
تقويم	يقوم المعلومات الصحية وأساليب الدعاية عنها (10%)	2=2.4 أو 3	2=2.4 أو 3	1= 1.2	6
تطبيق	يطبق المبادىء والتعميمات في مواقف جديدة (10 %)	2=2.4 أو 3	2=2.4 أو 3	1= 1.2	6
مجموع الفقرات		24	24	12	60

يتضح من هذا الجدول أنه يجوز التقريب عند حساب عدد الفقرات في كل خلية من خلايا الجدول. كما أنه ليس شرطا أن تملأ الفقرات كافة خلايا الجدول. وكذلك فإنه يمكن الاكتفاء بمستويات هدفية معينة دون غيرها على أن يتم تبرير ذلك.

من هنا يستنتج بأن هناك نماذج متنوعة لتكوين جداول المواصفات للاختبارات ذات أقصى الأداء كأدوات قياس التحصيل والقدرات العقلية والاستعدادات. أما عند بناء مقاييس الاتجاهات أو الميول وغيرها، فإن مصمم أدوات القياس يستخدم جدول مواصفات يتضمن سلوكات من المجال

الانفعالي (الوجداني) كما وردت عند كراثول وبلوم وماسيا (Krathwol, Bloom, and Masia, 1964).

4 – تصميم وبناء عبارات أو فقرات الأداة

أشار لندكويست (Lindquist, 1936) إلى أن هناك سؤالين مهمين يجب على مطوري المقاييس النظر إليهما بعناية منذ الشروع بتصميم مقاييسهم هما:

- ما الذي سيتم قياسه ؟
- كيف يمكن تنفيذ عملية القياس ؟

ولأغراض بناء الفقرات فإن السؤال الثاني هو موضع الاهتمام، والذي بدوره يتطلب الاهتمام بالنشاطات التالية:

- **اختيار الشكل المناسب للفقرات Item formats**
- **التأكد من فاعلية نوع الفقرات وملاءمتها للفئة المستهدفة**
- **اختيار وتدريب القائمين على كتابة الفقرات Item writers**
- **كتابة الفقرات Writing the items**
- **ضبط كتابة الفقرات ونوعيتها Monitoring the quality**

إن اختيار شكل فقرات الإختبار (الانتقائية أو الصياغية) ربما يكون قد حسم من خلال ما تم تحديده في جدول المواصفات، لأن غير ذلك يجب تحديده بداية هذه المرحلة وليس أن يترك لمزاج من سيكتب الفقرة. وعليه فإن: (1) مراجعة للمقاييس المشابهة والمنشورة، و (2) استشارة الخبراء في موضوع المقياس موضع الاهتمام يفيد في عملية اختيار شكل الفقرات.

أما القرار المتعلق بكون فقرات المقياس مفتوحة النهاية أو غير ذلك فإنه يرتبط بمدى توفر إمكانيات التصحيح ونوع المعلومات المراد جمعها عن المستهدفين، وكذلك يتوقف على الفئة العمرية لهم وغيرها من العوامل.

بعد ذلك لا بد من تحديد معايير **Standards** ومواصفات **Specifications** الفقرات المطلوبة، إضافة إلى موجهات لكتاب الفقرات، وخاصة إذا كان هؤلاء مبتدئين أو غير محترفين في هذا العمل.

وعلى أية حال، فإنه لا يوجد نوع من الفقرات يفضل على النوع الآخر بشكل مطلق. بل إن الاختيار يتم في ضوء استعراض عوامل قوة أو ضعف كل نوع منها. فالفقرات الانتقائية تتميز بقدرتها على تغطية أجزاء من المجالات والمحتويات العلمية الدراسية أكبر من قدرة الفقرات المقالية، وتعطي مؤشرات أكثر ثباتا واستقرارا، إضافة إلى سهولة تصحيحها يدويا أو آليا. كما أنها تساعد في عمليات تحديد المطلوب من السؤال بشكل أكثر دقة، وتستبعد أية تدخل ذاتي للمصحح ولذلك فهي تعرف بالأسئلة الموضوعية.

أما الفقرات الصياغية، فلا تحتاج زمنا طويلا لإعدادها، وهي أسهل صياغة على كل الأحوال ، وبنفس الوقت فهي بحاجة إلى وقت أطول في التصحيح وقد تتفاوت درجات الطلبة على السؤال الواحد نظرا لاختلاف المصحح، وربما يكون ذلك بسبب ما يعرف بتأثير عامل التوريه **pluffing**، أي محاولة بعض الممتحنين الدوران حول الإجابة بأسلوب كتابي وإنشائي يوهم المصحح بأنه طالب متمكن ويعرف الجواب، ويساعده في ذلك أسلوبه المميز في الكتابة والتعبير. ولا يخفى أن أكبر عيب يواجه الفقرات الانتقائية هو احتمال أن يجيب الممتحن عليها بفعل التخمين العشوائي. كما أن للفقرات الانتقائية قدرة جيدة على قياس المستويات العقلية بتنوع

وفعالية، إلا أن الفقرات المقالية أكثر قدرة في قياس المستويات العقلية المتقدمة. إضافة إلى أنها توفر فرصة للممتحن أن يظهر ما لديه من إبداع وحسن تعبير عن نفسه وقدراته بشكل أفضل.

ولاشك أن عددا من الأمور يجب اعتبارها عند اختيار نوع من الفقرات دون آخر ، فعندما يكون عدد الممتحنين كبيرا جدا فإن النصيحة أن يكون الخيار أسئلة موضوعية ولكن ما يقلل من هذا الاتجاه إذا كان الحيز الذي سيجلس فيه الممتحنون (قاعة الامتحان) ضيقا، الأمر الذي يجعلهم يجلسون ملتصقين بجوار بعضهم، وهذا يجعل إدارة الإختبار أمرا أكثر صعوبة. ويلخص الجدول التالي مقارنة أولية بين كل من الفقرات الموضوعية (الانتقائية) والفقرات الصياغية (المقالية وفقرات الإجابة القصيرة والتكميل) بدلالة عدد من العوامل:

مقارنة بين أنواع فقرات الإختبارات في ضوء عدد من العوامل

الرقم	العامل	نوع الفقرة		
		صياغية		انتقائية
		مقالية	إجابة قصيرة وتكميل	
1	قياس القدرة على حل مسألة جوهرية	+ +	+	+ +
2	قياس القدرة على التنظيم والتكامل والتركيب	+ +	+	– –
3	قياس الأصالة والإبداع في تناول المشكلات	+ +	+	– –
4	لها قيمة مميزة في التشخيص	++	+	+
5	إمكانية أن تتضمن عينة ممثلة لأهداف التدريس بشكل متوازن	– –	–	+ +
6	إمكانية تمثيل المحتوى الدراسي بشكل متوازن	– –	–	+ +
7	ضآلة فرص تخمين الإجابة	+ +	+ +	– –
8	ضآلة تأثر درجة المفحوص بعامل التورية	– –	–	+ +
9	إمكانية الثبات في تقديرات المصححين	– –	–	+ +
10	دقيقة في التمييز بين مستويات الأداء للمفحوصين	– –	–	+ +
11	إمكانية التصحيح آليا أو من قبل أناس غير محترفين	– –	–	+ +
12	السرعة في التصحيح	– –	–	+ +
13	تأخذ وقتا قصيرا في كتابة الفقرات	+	+	– –

تم اقتباس معظم بنود الجدول من (Thorndike and Hagen, 1986) وتشير الرموز (++) إلى أقصى فائدة، (+) إلى فائدة عادية، (––) إلى عديمة الفائدة، (–) إلى قليلة الفائدة.

وجدير بالذكر، فإن أي من أشكال الفقرات المذكورة يمكن أن يستخدم في بناء أدوات قياس الجوانب المعرفية أو ما يعرف بمقاييس أقصى الأداء

كما في اختبارات التحصيل والاستعداد والقدرات، مع الأخذ بعين الاعتبار عوامل تفضيل أي منهـا علـى الأخـرى (**Brwon, 1972; Mehrens and Lehman, 1991; Nitko, 1983; Popham, 1981; Sax, 1980**). بينمـا يـتم اختيار أشكال أخرى من الفقرات عند الشروع ببناء المقـاييس المتعلقـة بالجوانـب غيـر المعرفيـة (الميـول **interests** والاتجاهـات **attitudes** والشخـصية **personality** وغيرهـا) كمـا فـي الإسـتبانات والقـوائم **Inventories**. وقد تعددت أساليب بناء مقاييس الجوانـب غيـر المعرفيـة وتنوعت، فهناك أكثر من خمسة أساليب لقياس الاتجـاه، ومثلهـا لقيـاس الميول، أما قياس الشخصية فيبدو أكثر تفصيلا وتعقيدا من تلك. وربما سيتم تناول هذا المجال في كتب أخرى أو في أجزاء خاصة من هذا الكتاب.

5- مراجعة فقرات المقياس

بعد كتابة عبارات أداة القياس بصورة أولية، يقوم مصمم تلك الأداة بعرضها بشكل ودي (غير رسمي) على مجموعة من المهتمين من الزملاء والخبراء في موضوع الأداة بشرط توفر الكفاءة والرغبة لديهم في التعاون لفحص كل من:

- اللغة ومفرداتها **wording**،
- الدقة والصحة **accuracy and precision**،
- درجة غموض المعنى أو وضوحه **ambiguity**،
- قواعد اللغة **grammar**.

وقد يؤدي ذلك الفحص إلى اكتشاف نوعية الفقـرات والتعـرف علـى جوانب الضعف والقوة في كل منها. الأمر الذي يستوجب تبعا لذلك إجراء التعديل المناسب لتطويرها على نحو أفضل. كما يتبع ذلك كلـه مراجعـة الفقرات بأسلوب تفصيلي ورسمي **formal** يتضمن فحص الفقرات بـشكل أدق من الفحص السابق ويكون عندها التركيز على:

- دقة الفقرة وصحة المعلومات المتضمنة فيها **accuracy and precision**
- مدى ملاءمة الفقرة لجدول المواصفات **relevance to test specification**
- مدى تحيز الفقرة **appearance of bias**
- مستوى مقروئية الفقرة **level of readability**
- أخطاء في التصميم **construction flaws**

ومن المفروض أن يضم فريق المراجعين عددا من المتخصصين في مادة الإختبارات موضع البحث للتأكد من صحة ودقة المعلومات التي تتضمنها فقرات المقياس، ومن متخصصين في القياس والإختبارات، للتأكد من أن صياغة الفقرات تراعي قواعد الصياغة، وتقلل ما أمكن من العيوب الشائعة في صياغة كل نوع من أنواع الفقرات. كذلك لا بد أن يتضمن الفريق عددا من المهتمين في اللغويات لضمان سلامة الصياغة اللغوية. ولا يخفى أن يكون مراجع الإختبار على دراية بخصائص الأفراد المستهدفين بالمقياس من ناحية العمر والمستوى الثقافي والعلمي والاجتماعي لمراعاة عدم إظهار الفقرات أي نوع من التحيز تجاه عوامل معينة كالعمر أو الثقافة أو الجنس وغيرها.

والمعروف أيضا أنه نتيجة لعمليات المراجعة فإن عددا من الفقرات قد يتم شطبه أو حذفه من المقياس، أو قد يصار إلى تطويره وتعديله بالشكل المطلوب. أو ربما إعادة تصنيف بعض من الفقرات في أبعاد جديدة يتألف منها المقياس وتحت عناوين جديدة.

6- التطبيق الأولي لفقرات المقياس Preliminary Test Items Tryout

تأتي هذه الخطوة قبيل طباعة المقياس وإخراجه بصورته النهائية ليكون جاهزا للتطبيق والاستخدام العملي والميداني. إذ ينصح بتطبيق

فقرات المقياس على عينة صغيرة من المعنيين (20-30 شخصا)، وربما يكون الرقم المطلوب أكبر من ذلك (100 – 200) إذا كانت النية تتجه إلى استخدام الإختبار على مدى أوسع مستقبلا أو في أغراض تجارية مثلا. وتساعد هذه المرحلة من التطبيق في الحصول على:

خصائص الفقرات كما في معاملات الشدة أو الشيوع والصعوبة والتمييز،

- **تحديد الزمن اللازم** لإكمال الإجابة على الإختبار كاملا، أو على كل فقرة،
- **رصد ردود أفعال المستجيبين** على فقرات المقياس وأبعاده أثناء التطبيق.

ولمزيد من الفائدة، فإن فحص توزيع درجات المفحوصين على كل فقرة وعلى المقياس ككل يساعد في دراسة مدى صعوبة تلك الفقرة ومدى ملاءمتها لمستوى المفحوصين بشكل أفضل. كما وينصح كذلك بعقد مؤتمر مصغر يضم النسبة الأكبر من المفحوصين بعد هذه المرحلة من أجل استطلاع اقتراحاتهم على فقرات المقياس من أجل تطويرها على نحو أفضل. وبالنهاية، فإن صيغة المقياس بصورته النهائية ربما تكون جاهزة في ضوء النتائج والدروس المتعلمة بعد هذه المرحلة من التطبيق.

7- الخطوات القادمة The next steps

يعد المقياس صالحا للاستخدام في الواقع الميداني إذا خضع لكافة الخطوات الست السابقة، وهذا لا يعني بالضرورة أن يكون المقياس كاملا ومثاليا، بل إن عمليات التطوير والتحسين لن تتوقف. فعلاوة على ضرورة كتابة تعليمات تخص عمليات تطبيق المقياس وأخرى تتعلق بتصحيحه، فإن

هناك مجموعة من الإجراءات الكمية والإحصائية التي يجب القيام بها لضمان اعتبار المقياس أداة قياس فعالة إلى درجة كبيرة وهي:

- **تحليل الفقرات Item analysis**
- **دراسات الصدق Validation studies**
- **دراسات الثبات Reliability studies**
- **اشتقاق معايير Setting standards**
- **استراتيجيات التصحيح Scoring strategies** ويتم بحث كيفية التصحيح آليا أو يدويا ويتوقف هذا على شكل فقرات المقياس، كما يتطلب ذلك التعرف على التخمين وكيفية التحكم به وضبط أثره منهجيا.

وسيرد شرح مفصل عن هذه الدراسات في فصول لاحقة. مبينا بأن هذه الإجراءات قد تختلف باختلاف نوع المقياس عندما تصنف حسب طريقة تفسير النتائج إلى معيارية المرجع أو محكية المرجع.

فالإختبار معياري المرجع Norm-Referenced-Test هو الذي يتم تقييم مستوى أداء الفرد عليه بالمقارنة مع أداء مجموعة أكبر تدعى (مجموعة معيارية)، قد تكون عينة وطنية تمثل امتدادا عريضا ومتنوعا من الأفراد. وعادة ما يتم مقارنة أداء الطلبة أو مستوى المدارس أو حتى المناطق التعليمية مع متوسط أداء المجموعة المعيارية. والهدف الأساس من هذا النوع من التقييم هو ترتيب الطلبة وليس قياس تحصيلهم تجاه بعض معايير الأداء.

أما **الإختبار محكي المرجع Criterion-Referenced-Test** فهو الذي يتم تقييم مستوى الأداء عليه من خلال مقارنة ذلك المستوى بهدف تعلمي أو بأداء محكي محدد، وليس مقارنته بمستوى أداء الأفراد الآخرين.

ويساعد هذا النوع من التقييم في تحديد مستوى أداء الطلبة على أهداف ومستويات أكثر من مقارنة أدائهم مع أداء مجموعة معيارية محلية أو وطنية. وقد يصل جميع الطلبة أو ربما لا يصل أحد منهم مستوى الأداء المحكي.

قواعد عامة لكتابة فقرات الإختبار

I. الفقرات الانتقائية

هناك عدد من الفقرات التي تندرج تحت هذا النوع، مثل فقرات الاختيار من متعدد، فقرات المزاوجة، فقرات البديلين أو الصواب والخطأ. وسيتم التعرض هنا إلى مناقشة شروط بناء فقرات الاختيار من متعدد وفقرات المزاوجة فقط. على اعتبار أن فقرات الخطأ والصواب ما هي إلا حالة خاصة من فقرات الاختيار من متعدد.

I. 1: فقرات الاختيار من متعدد Multiple Choice Items

تتألف الفقرة الموضوعية من نوع الاختيار من متعدد من: (1) متن أو مقدمة الفقرة **Item Stem**، (2) بدائل الفقرة **Alternatives** والتي تتضمن الجواب الصحيح والمموهات **Distractors**.

والأصل في الفقرة الجيدة أن ترتبط بتحقيق هدف **objective** معين يرمي الإختبار إلى قياسه، كما يجب أن ترتبط بمحتوى دراسي **content** أو بقدرة **ability** أو استعداد **aptitude** معين.

المموه الجيد هو بديل جذاب يختاره نسبة معينة من الممتحنين. بحيث تكون نسبة من يختاره من مجموعة الأداء المنخفض أكبر من نسبة من يختاره ممن ينتمي إلى فئة الأداء المرتفع. بذلك يجب **استبعاد أو تعديل أو تطوير** أي مموه لا يختاره أي من الممتحنين ضعيفي المستوى. ولحل هذا الإشكال بدرجة معقولة، ينصح بطرح الفقرة أولا على أنها سؤال إنشائي (مفتوح) على مجموعة من الممتحنين، ثم يصار إلى اشتقاق المموهات وتطويرها من خلال حصر وتحليل أخطاء هؤلاء الممتحنين.

والجدير بالذكر أن تحليل جاذبية المموهات يتم تناوله بشكل مفصل خلال الفصل الرابع من هذا الكتاب. إذ يكون المموه جيدا (جذابا) لفئة ضعاف مستوى الأداء إذا تمتع بمعامل جاذبية سالب.

عدد المموهات

يمكن القول بأنه كلما زاد عدد البدائل في الفقرة تقل فرصة نجاح الممتحن بفعل التخمين. لذلك فإن من الأفضل إضافة بدائل أكثر، شريطة أن تكون ذات معنى من حيث صلتها بالمجال السلوكي أو المحتوى الدراسي من ناحية، وأن تتوفر فيها جاذبية معينة يختارها الممتحنون ممن درجاتهم الكلية على الإختبار منخفضة من ناحية أخرى.

والجدير بالذكر أن فرصة الحصول على الإجابة الصحيحة بفعل التخمين تزداد كلما قل عدد المموهات. فهي تساوي 50 % في حالة الفقرة ببديلين (فقرة الصواب والخطأ)، وهي 25 % في حالة الفقرة ذات الأربعة بدائل، كما تصل الفرصة 20 % في حالة الفقرة ذات الخمسة بدائل، وهكذا.

***. استخدام لغة واضحة وسهلة في متن الفقرة تراعي الاختصار والوضوح**

لا تستوجب الإجابة على الفقرة التي تتمتع بصياغة جيدة قراءة كافة البدائل، من أجل اختيار الجواب الصحيح، إذ أن المفروض أن يتضمن المتن كافة المعلومات التي تتضمنها الفقرة. من ناحية ثانية، قد يلجأ بعض مطوري الإختبارات إلى اختزال المعلومات في المتن أو المبالغة فيها بهدف تضليل الممتحن، الأمر الذي يتنافى مع أهداف الامتحان كوسيلة تربوية لتقويم أدائه. فالغموض أو التضليل ليس هدفا بحد ذاته، وليس غرضا من أغراض القياس. والمثالان التاليان يحوي كل منهما فقرة، متنها غير واضح، أو لا يتضمن السؤال المطلوب فعلا:

مثال (1)

صياغة مقترحة	صياغة ضعيفة
أبو ظبي عاصمة دولة ----	أبو ظبي ----
أ) البحرين	أ) مدينة ساحلية
ب) الإمارات	ب) شبه جزيرة
ج) قطر	ج) عاصمة دولة الإمارات
د) الكويت	د) مدينة خليجية

نلاحظ أن الصياغة الأولى للفقرة في هذا المثال ضعيفة، بحيث أنها لا تحوي السؤال المطلوب الإجابة عنه. هذا بالرغم من أن كافة البدائل الواردة في تلك الفقرة صحيحة، كونها صفات مدينة أبو ظبي فعلا. لذلك، بات من الضروري أن يصاغ متن الفقرة على النحو المقترح (في الصياغة الأفضل) أو تحويلها إلى فقرة مقالية على النحو التالي: **ما عاصمة دولة الإمارات العربية المتحدة ؟**

مثال (2)

صياغة مقترحة	صياغة ضعيفة
تفرز خلايا جزر البنكرياس مادة ---.	خلايا جزر البنكرياس ----.
أ) تريبسين	أ) تحتوي على قنوات
ب) أنسولين	ب) تختفي مع التقدم في العمر
ج) سكرتين	ج) تقع على أطراف البنكرياس
د) أدرينالين	د) تنتج الأنسولين

تم اقتباس هذه الفقرة من ثورانـدايك و هـاجن (**Thorndike and Hagen, 1986**). وفي الصياغة المقترحة، تم تعديل البـدائل والمـتن. لأن السؤال المطلوب في الفقرة غير واضح. على الرغم من أن البدائل جميعها

ذات علاقة معينة بجزر البنكرياس. وقد تم تعديل الصياغة بحيث أصبح متن الفقرة أكثر وضوحا، لأنه تضمن السؤال المطلوب بوضوح. وتم كذلك إعادة تنظيم البدائل (المموهات) بحيث أصبحت أكثر جاذبية.

***. تجنب وجود ترابط لفظي بين المتن والبدائل (اقتران تكراري)**

كأن تظهر كلمة في المتن وتتكرر في أحد البدائل. الأمر الذي يشير إلى موقع الإجابة الصحيحة. وهذا يؤدي بدوره إلى إضعاف الفقرة وجعل الجواب المطلوب مكشوفا. ويتضح ذلك من عرض المثالين التاليين:

مثال (1)

صياغة ضعيفة	صياغة مقترحة
تقوم وزارة التربية والتعليم على رعاية الشؤون ----. **أ) الاقتصادية** **ب) التعليمية** **ج) الأمنية** **د) الاجتماعية**	**أهم وظيفة تقوم بها وزارة التربية والتعليم هي ----.** **أ) تطوير المستوى النوعي للإنسان** **ب) إعداد مناهج تعليمية معاصرة** **ج) إعداد امتحانات تحصيلية** **د) توفير فرص تعليمية للطلبة**

كان من الأفضل أن تصاغ هذه الفقرة كما ظهرت في الصياغة المقترحة، أو أن تكتب على هيئة سؤال مقالي (إنشائي) مثل: **(ما الوظائف والخدمات التي تقوم بها وزارة التربية والتعليم ؟).**

مثال (2)

صياغة مقترحة	صياغة ضعيفة
ما نوع الصدق المطلوب التحقق منه في حالة الإختبارات التي تمثل منهاجا دراسيا أو مادة تعليمية؟ أ) التنبؤ ب) البناء ج) التزامني د) المحتوى	لمعرفة مدى تغطية اختبار لمحتوى دراسي معين، لا بد من التحقق من صدق---. أ) التنبؤ ب) البناء ج) التزامني د) المحتوى

في الصياغة الضعيفة يتم استبعاد البديل (ج) تلقائيا. بسبب التكملة التلقائية لكلمة (صدق) التي تستوجب وجود كلمة تبدأ باللام القمرية. وعندها ينحصر الجواب المطلوب من بين البدائل الأخرى (أ) و (ب) و (د).

***.تجنب التداخل بين بدائل الفقرة**

يتضمن البديل الواحد أحيانا بديلا آخر أو أكثر. وبالتالي تفقد مموهات الفقرة جاذبيتها في أن يختارها الممتحن. بمعنى آخر، لن يكون لتلك الفقرة إجابة صحيحة واحدة كما في المثال التالي:

صياغة مقترحة	صياغة ضعيفة
عدد سكان المنامة يساوي ---	عدد سكان المنامة يساوي ---
أ) 80000	أ) 80000
ب) 60000	ب) أكثر من 60000
ج) 50000	ج) أكثر من 50000
د) 70000	د) بين 40000 و 70000

يلاحظ في الصياغة الضعيفة أن البدائل (أ) و (ب) و (ج) متداخلة ومتعادلة تماما. إذ أن البديل (ب) يتضمن البديلين (أ) و (ج). ويقع البديل (ب) ضمن حدي البديل (د). وكذلك كان البديل (ج) يتضمن البديل (أ). أما الصياغة المقترحة، فقد أظهرت إخراجا أفضل للبدائل، جعلها مستقلة عن بعض من ناحية. وأصبحت متشابهة في مدى جذبها للمستجيبين من ناحية أخرى.

***. تجنب التفاوت في طول البدائل**

فقد يلجأ بعض مطوري الإختبارات أحيانا إلى التفصيل في البديل الصحيح، فيظهر أطول من غيره، ويكون بالتالي بديلا جاذبا للمفحوص. الأمر الذي يجعل الفقرة ضعيفة ويستطيع الفرد إجابتها دون التمعن والتدقيق في محتواها كما في المثال التالي:

صياغة ضعيفة	صياغة مقترحة
أهم صفات الإختبار الجيد ---	أهم صفات الإختبار الجيد ---
أ) قياس الأهداف التي بني من أجل تحقيقها كاختبار الرياضيات للصف السادس الابتدائي الذي يغطي خمس وحدات دراسية في مجال العمليات الحسابية والكسور والجذور التربيعية والمساحات والحجوم بواقع نسبة مئوية معينة لكل وحدة.	أ) الصدق
ب) الثبات	ب) الثبات
ج) الموضوعية	ج) الموضوعية
د) سهولة التصحيح	د) سهولة التصحيح

فالبديل رقم (أ) في الصياغة الضعيفة يوحي بأنه الأكثر صحة وجاذبية من قبل معظم الممتحنين. وفي هذه الحالة ينصح باستبدال نص هذا البديل بما يدل عليه فعلا وهو صفة الصدق.

***. تجنب المنبهات clues القواعدية والنحوية في المتن**

تكون المنبهات لغوية قواعدية أو لفظية بصرف النظر عن اللغة المستخدمة بالسؤال. الأمر الذي لو احتوت الفقرة بعضا منها لكان ذلك مؤشرا على ضعفها. كون ذلك يساعد المستجيب على اختيار الجواب الصحيح دون بذل الجهد المطلوب كما في المثالين التاليين:

مثال (1)

صياغة ضعيفة	صياغة مقترحة
الذي هاجر مع الرسول (ص) من مكة المكرمة إلى المدينة المنورة هو --.	**الذي هاجر مع الرسول (ص) من مكة المكرمة إلى المدينة المنورة (هو)/(هي)--.**
أ) عمر بن الخطاب	**أ) عمر بن الخطاب**
ب) أسماء بنت أبي بكر	**ب) أسماء بنت أبي بكر**
ج) أبو بكر الصديق	**ج) أبو بكر الصديق**
د) خديجة بنت خويلد	**د) خديجة بنت خويلد**

في الصياغة الضعيفة، يستبعد المفحوص البديلين (ب) و (د) لأنهما نساء، ولكون متن الفقرة يشير إلى أن المطلوب هو أحد البديلين (أ) أو (ج)، نظرا لأنه تمت الإشارة في المتن إلى ذلك من خلال الضمير (هو). في حين جاءت الصياغة المقترحة بحيث تكون البدائل جميعها متكافئة. كما أنه من الممكن أن تصاغ الفقرة بشكل إنشائي مثل:

"من أول من هاجر مع الرسول (ص) من مكة المكرمة إلى المدينة المنورة ؟ " .

مثال (2)

صياغة مقترحة	صياغة ضعيفة
Every day a healthy boy eats a/an - --- a. cheese b. apple c. loaf d. orange or Every day a healthy boy eats ----- a. cheese b. apple c. loaf d. orange	Every day a healthy boy eats an ---- a. cheese b. apple c. loaf d. orange

في الصياغة الضعيفة، يقوم الممتحن باستبعاد البديلين (a) و (c) لأنهما غير متسقين من ناحية قواعدية مع المتن لوجود (an). الأمر، الذي تم تجنب ذلك في الصياغة المقترحة وجعل البدائل متجانسة ومتعادلة من حيث جاذبيتها.

***. تجنب أن تتكرر كلمة أو عبارة في بدائل الفقرة**

إذا تكررت كلمة أو عبارة في البدائل أو في معظمها، فإنه ينصح بنقل الكلمة أو العبارة المكررة إلى متن الفقرة، لأن غير ذلك يجعل إخراج الفقرة ركيكا كما في المثال التالي:

صياغة ضعيفة	صياغة مقترحة
أهم خواص عينة الدراسة ---. أ) أن تكون العينة كبيرة. ب) أن تكون العينة ممثلة للمجتمع. ج) أن تكون العينة عشوائية. د) أن تكـــون العينـــة طبقيـــة أو عنقودية.	أهم خواص عينة الدراســة، أن تكــون العينة---. أ) كبيرة. ب) ممثلة للمجتمع. ج) عشوائية. د) طبقية أو عنقودية.

*. تجنب أن تكون البدائل متباينة (غير متجانسة)

يلجأ المفحوص إلى استبعاد أي من البدائل التـــي لا تتـــسجم مـــع موضوع الفقرة. لذلك، بات من الضروري أن تكون بدائل الفقرة متقاربـــة ومتجانسة ما أمكن، إذا أن صعوبة الفقرة تزداد بزيادة تجانس بدائلها كمـــا في المثالين التاليين:

مثال (1)

صياغة ضعيفة	صياغة مقترحة
ما أطول نهر في أمريكا ؟	ما أطول نهر في أمريكا ؟
أ) الأمازون	أ) الأمازون
ب) المسيسبي	ب) المسيسبي
ج) بوتوماك	ج) بوتوماك
د) إيفل	د) أوهايو

مثال (2)

صياغة ضعيفة	صياغة مقترحة
ما عاصمة هولندا ؟ أ) أمستردام ب) النيل ج) أستوكهولم د) أوسلو	ما عاصمة هولندا ؟ أ) أمستردام ب) كوبنهاجن ج) أستوكهولم د) أوسلو

للإجابة على أي من هذين المثالين يتم استبعاد كل من البديل (د) من المثال (1) لأنه ليس نهرا، وكذلك البديل (ب) من المثال (2) لأنه ليس عاصمة دولة اسكندنافية، وإنما هو اسم أحد الأنهار. الأمر الذي جعل الصياغة المقترحة تتلافى هذا العيب، وبالتالي ظهرت الفقرة بصورة أفضل. بحيث كانت البدائل متجانسة وتشير بمجملها إلى مدن أو عواصم بلدان اسكندنافية.

***. تجنب وجود كلمات تفيد الجزم والإطلاق أو التعميم والتخصيص**

مثل استخدام كلمات مثل: دائما، فقط، كل، جميع، إطلاقا، حتما، قطعا، always ، all، absolutely لأن هذه الكلمات تشير في الغالب إلى الإجابة بالنفي كما في المثال التالي:

صياغة ضعيفة	صياغة مقترحة
الخضار دائما أكثر فائدة من البروتينات (نعم/لا)	الخضار أفضل من البر وتينات (نعم/لا)
جميع أنواع البكتيريا تسبب الأمراض (نعم/لا)	جميع أنواع البكتيريا التي تسبب الأمراض هي طفيليات (نعم/لا)

يتضح بأن الإجابة على الفقرتين الأوليتين (في حالة الصياغة الضعيفة) تكون بالنفي. لذلك تم اقتراح الصياغة المقابلة التي قد تحتمل النفي (لا) أو الإثبات (نعم) وليس النفي فقط.

***. تجنب استخدام الكلمات الاحتمالية**

إن استخدام الكلمات مثل: (ربما)، (أحيانا)، (بعض الأحيان)، (معظم الحالات)، (إلى حد ما) تقود في الغالب لأن تكون الإجابات بنعم أو بالموافقة كما في المثال التالي.

صياغة ضعيفة	صياغة مقترحة
تكون الخضار أفضل من البروتينيات أحيانا (نعم/لا)	الخضار أفضل من البروتينات (نعم/لا)
يستخدم التجفيف كثيرا في حفظ الأطعمة (نعم/لا) أو يستخدم التجفيف أحيانا في حفظ الأطعمة (نعم/لا)	يمكن حفظ الفاكهة بالتجفيف (نعم/لا)

***. تجنب استخدام البدائل مثل: (كل ما سبق)، (جميع ما ذكر)، (لا شيء مما ذكر)**

إن اختيار أو استبعاد مثل هذه البدائل مرتبط باختيار أو استبعاد البدائل الأخرى. فعندما يكون البديلان الأول والثاني صحيحين، فهنا يحتمل أن يختار الممتحن البديل (كل ما سبق) بصرف النظر عن البدائل الأخرى. إذ لا يحتاج الممتحن في الغالب أن يتمعن بها أو حتى يقرأها. وبنفس الطريقة قد يتم اختيار البديل (لا شيء مما ذكر) إذا كان البديلان الأول والثاني خاطئين.

وبشكل عام يجب الحذر عند استخدام مثل هذه البدائل. أما إذا كان لا بد من استخدامها، فيجب أن يكون بعضها صحيحا و بعضها خاطئا. ويبين المثالان التاليان نموذجا لهذا الإرشاد.

مثال (1)

صياغة ضعيفة	صياغة مقترحة
يهدف البحث التجريبي بشكل رئيس إلى ---	يهدف البحث التجريبي بشكل رئيس إلى ---
أ) استقصاء علاقة السبب تيجة	أ) استقصاء علاقة السبب بالنتيجة
ب) المقارنة بين المجموعات ريبية	ب) المقارنة بين المجموعات ريبية
ج) دراسة العلاقات بين المتغيرات	ج) دراسة العلاقات بين المتغيرات
د) جميع ما ذكر	د) تقييم فعالية طرق المعاينة

هنا قد لا يكون الممتحن عارفا بشكل كامل بأن دراسة العلاقات بين المتغيرات ليست أحد أهداف البحث التجريبي كما هو بالنسبة للبديلين (أ) و (ب). وبالتالي، فهو يختار البديل (د) معتقدا بأنه هو البديل الصحيح.

مثال (2)

صياغة ضعيفة	صياغة مقترحة
أي مما يلي يعتبر من مقاييس التشتت؟	أي من التالية لا يعتبر من مقاييس التشتت؟
أ) الانحراف المعياري	أ) الانحراف المعياري
ب) المدى	ب) المنوال
ج) المنوال	ج) نصف المدى الربعي
د) جميع ما ذكر	د) التباين

في هذا المثال، يعتقد المفحوص بدرجة عالية بأن نصف المدى الربعي هو أحد مقاييس التشتت وبالتالي فهو يختار البديل (د) معتقدا بأنه هو البديل الصحيح ويحصل على الدرجة.

***. تجنب صياغة النفي والنفي المركب في متن الفقرة**

يفضل صياغة متن الفقرة في الاتجاه الإيجابي. وإذا كان لا بد من استخدام النفي، فيجب إبرازه لتنبيه الممتحن إليه بوضع خط تحته أو طباعته بخط كبير مميز كما في المثالين التاليين:

مثال (1)

صياغة ضعيفة	صياغة مقترحة
ليس أي من اختبارات الذكاء لا يحتاج إلى تقنين.	<u>ليس</u> أي من اختبارات الذكاء <u>لا</u> يحتاج إلى تقنين أو اختبارات الذكاء تحتاج إلى تقنين (نعم /لا)

في هذا المثال نرى بديلين للصياغة الضعيفة. كانت الأولى الإبقاء على النفي والنفي المركب وإبرازهما بوضع خط تحت كل منهما. بينما كانت الصياغة الثانية تحويل الصياغة إلى صياغة إثبات.

مثال (2)

صياغة ضعيفة	صياغة مقترحة
أي مما يلي لا يعتبر من مقاييس تت ؟ أ) المدى ب) المنوال ج) التباين د) الانحراف المعياري	مما يلي يعتبر من مقاييس تت ؟ أ) الوسيط ب) المنوال ج) التباين د) فاي

في هذا المثال، ظهرت الصياغة المقترحة بصيغة الإثبات عوضا عن صياغة النفي.

***. تجنب الإسهاب wording في متن الفقرة**

المفروض أن يكون متن الفقرة واضحا وقصيرا ما أمكن مع ضرورة الحفاظ على كمال المعنى. وبعكس ذلك فإن الاستفحال في مقدمة الفقرة ربما يحوي معلومات كثيرة ليس لها علاقة بالسؤال مباشرة. الأمر الذي قد يؤدي إلى تشتيت انتباه المفحوص وإضاعة وقته من ناحية، أو يجعل الفقرة تقيس أهدافا غير التي بنيت من أجلها.

صياغة ضعيفة	صياغة مقترحة
من المعروف أن الصدق أهم صفة يتمتع بها الإختبار الجيد، إذ يحتل الصدق الصفة رقم (1) في خصائص الإختبارات بأنواعها. كما يمكن تعريفه بالدرجة التي يمكن لأداة القياس أن تقيس ما وضعت من أجله. ومعروف أيضا أن الحديث عن صدق الإختبار يرتبط بشكل مباشر بأهداف الإختبار، فهناك الإختبار الموجه نحو إجراءات قبول الطلبة في برامج معينة مستقبلا، ويمكن أن يتم دراسة صدق المحك التنبؤي له. ما الصدق الذي يتم دراسته في حالة الإختبارات التحصيلية؟ **أ) المحتوى ب) التنبؤي** **ج) التلازمي د) البناء**	**ما الصدق الذي يجب دراسته عند تطوير اختبارات التحصيل؟** **أ) المحتوى** **ب) التنبؤي** **ج) التلازمي** **د) البناء**

لم تكن المقدمة الطويلة لتلك الفقرة ضرورية، طالما أنه كان بالإمكان توجيه السؤال المطلوب مباشرة ، ويمكن قياس الهدف المطلوب بفاعلية. هذا علاوة على أن قراءة مثل هذه الفقرة يستغرق وقتا أطول. الأمر الذي ربما يؤثر سلبا على الوقت المخصص لإجابة الفقرات الأخرى في الإختبار.

***. تجنب أن تعتمد إجابة فقرة على فقرة أخرى**

أي أن المفروض أن تكون الإجابة على فقرة ما غير مرتبطة بالإجابة على فقرة أخرى. بمعنى أن تكون إجابة الفقرات مستقلة عن بعضها البعض. بهذا يتم تجنب تكرار الخطأ وجعل العقوبة مزدوجة أو تراكمية. والمثالان التاليان يلقيا الضوء على هذه الحالة.

مثال (1)

صياغة ضعيفة	صياغة مقترحة
س10: ما حاصل ضرب الأعداد: 8، 9 ، 2 **س11: ما الجذر التربيعي لجواب المسألة التي وردت في السؤال (10)**	**س10: ما حاصل ضرب الأعداد: 8، 9 ، 2** **س11: ما الجذر التربيعي للعدد 144 ؟**

في الصياغة الضعيفة من هذا المثال، إذا كان جواب المسألة رقم (10) خطأ، فسيكون جواب المسألة رقم (11) خطأ أيضا، وفي بعض الأحيان يكون الجذر التربيعي المحسوب صحيحا حتى لو كان لجواب خاطئ أساسا. في مثل ذلك يتعرض المفحوص لعقوبة مزدوجة.

مثال (2)

صياغة ضعيفة	صياغة مقترحة
س5: ينتج مرض الأسقربوط عن نقص في فيتامين: أ) (A) ب) (B1) ج) (B12) د) (C)	س5: ينتج مرض الأسقربوط عن نقص في فيتامين: أ) (A) ب) (B1) ج) (B12) د) (C)
س6: ومن المصادر الجيدة لهذا الفيتامين: أ) عصير البرتقال ب) الأرز المقشور ج) زيت كبد الحوت د) الكبد	س6: ما المصدر الرئيس لفيتامين (C)؟ أ) عصير البرتقال ب) الأرز المقشور ج) زيت كبد الحوت د) الكبد

في الصياغة الضعيفة، إذا كان جواب الفقرة رقم (5) خطأ، فسيكون جواب الفقرة رقم (6) خطأ أيضا. في مثل ذلك يتعرض الممتحن لعقوبة مزدوجة. أما في الصياغة المقترحة ، يكون الأداء على الفقرة الواحدة مستقلا عن الأداء على الأخرى.

*. أن تتسق المموهات مع متن الفقرة علميا ومنطقيا

المفروض أن تكون المموهات متسقة منطقيا مع متن الفقرة ومماثلة للأخطاء الشائعة بين الممتحنين في مرحلة دراسية معينة كما في المثالين التاليين:

مثال (1)

صياغة ضعيفة	صياغة مقترحة
1- أي العناصر التالية نجده في البروتينات وليس في الكربوهيدرات أو الدهون ؟ أ) ثاني أكسيد الكربون ب) الأكسجين ج) الماء د) النيتروجين	2- أي المواد التالية نجده في البروتينات وليس في الكربوهيدرات أو الدهون ؟ أ) ثاني أكسيد الكربون ب) الأكسجين ج) الماء د) النيتروجين

في هذا المثال نرى بأن البديلين (أ) و (ج) هما مركبان وليسا عنصرين، ولذا فهما لا يتسقان مع المتن الذي يتضمن السؤال عن عناصر. لذا، فقد تم تصحيح الصياغة باستبدال كلمة العناصر بالمواد.

مثال (2)

صياغة ضعيفة	صياغة مقترحة
ما الغاز الذي يتكون في الخلايا بعد احتراق الطعام ويمر في الرئتين ثم يطرد إلى الخارج ؟ أ) ثاني أكسيد الكربون ب) الأكسجين ج) الهيليوم د) الكلور	ما الغاز الذي يتكون في الخلايا بعد احتراق الطعام ويمر في الرئتين ثم يطرد إلى الخارج ؟ أ) ثاني أكسيد الكربون ب) الأكسجين ج) الهيدروجين د) النيتروجين

في الصياغة الضعيفة من هذا المثال نرى بأن البديلين (ج) و(د) غير جذابين ومستبعدان تماما، لأنه من غير المنطق الاعتقاد بأن غازا مثل الهيليوم أو الكلور يمر في رئتي الإنسان بشكل طبيعي. ولذلك تم استبدالهما بغازين آخرين مثل الهيدروجين والنيتروجين.

***. يجب أن تتناول كل فقرة نصا مهما أو فكرة جوهرية**

المفروض أن تتضمن الفقرة معلومات مهمة، تساعد في تقييم هدف تعلمي مهما أيضا. إذ ليس من المعقول أن تكون الفقرة مجرد سؤال تعجيزي، لا يقيس هدفا قيما أو مخرجا تعليميا ذا قيمة، كما في المثال التالي.

صياغة ضعيفة	صياغة مقترحة
كان معدل الوفيات في جميع الحوادث نهاية العام الماضي في كل 100.000 من السكان من الفئة العمرية 15 – 24 يساوي ----	في عام 2000، كان السبب الرئيسي للوفيات بين أفراد الفئة العمرية 15-24 هو---
أ) 69.00 %	أ) الحوادث
ب) 69.10 %	ب) السرطان
ج) 69.01 %	ج) الأمراض الصدرية
د) 69.12 %	د) أمراض القلب

إن التعرف على أسباب وفيات فئة عمرية معينة في المجتمع (كما في الصياغة المقترحة) أكثر فائدة من التعرف على نسبة وفيات هذه الفئة (الصياغة الضعيفة). وعندما تكون النسب متقاربة جدا، لا يكون الفرق بينها على درجة كبيرة من الأهمية. كما أن التعرف على معدل الوفيات لأقرب مائة ليس على درجة من الأهمية أيضا مقارنة مع فحص معرفة أسباب الوفيات بالنسبة لفئة عمرية معينة.

***. يتم توزيع عدد الإجابات بالتساوي على البدائل**

ينصح بأن تكون مواقع الإجابات الصحيحة موزعا عشوائيا على بدائل الفقرات. لأنه يخشى أن يكتشف بعض الممتحنين ترتيبا معينا لمواقع إجابات الفقرات إذا تم ترتيبها بنسق منتظم. الأمر الذي يساعد الطلبة على اختيار الإجابة دون الاعتماد على ما لديه من معرفة علمية أصيلة أو متعلمة.

***. تجنب أن تتضمن الفقرة نصا حرفيا منقولا من الكتاب أو النص المقرر**

في مثل هذه الحالة، تعتمد الإجابة على الفقرة على ما يتذكره الطالب أثناء قراءته النص بالرغم من أن معنى هذا الجزء في الإختبار قد يختلف عنه في النص الوارد في الكتاب. والمطلوب هنا أن يستخدم واضع الإختبار ألفاظا وعبارات غير الواردة في الكتاب. ويبدو الأمر أكثر خطأ عندما يتم اقتباس عبارة بارزة كعنوان وحدة أو نص قانون أو غيره.

***. أن تكون الفقرات صحيحة من ناحية المحتوى**

من المفروض أن تتضمن الفقرة بديلا (صحيحا) واحدا أو البديل (الأكثر صحة) حسب طبيعة المطلوب من الفقرة. وهذا يستوجب تصميم مفتاح التصحيح بدقة كبيرة كي لا يطرأ خلل فاضح أثناء التصحيح، فينجح الراسب ويخفق الناجح. كما يجب الحذر بأن لا تحتمل الفقرة الواحدة أكثر من بديل صحيح. وهنا ينصح أن يقوم مطور الإختبار بتجريب فقرات اختباره على عينة مناسبة من الممتحنين في وقت مناسب.

***. ضرورة مراجعة الإختبار قبيل تطبيقه**

المفروض أن يتم مراجعة تعليمات الامتحان وعباراته من ناحية لغوية وعلمية قبل إخراجه بشكله النهائي. ذلك لأنه من المفروض أن يتم استغلال وقت الامتحان من قبل الممتحنين في التفكير والاستجابة فقط، وليس في تصحيح أخطاء لغوية، أو بتشتيت انتباه المفحوصين بتعليمات

جديدة. فقد يلجأ بعض المعلمين إلى الشروع في تصحيح الأخطاء المطبعية، وتوجيه الممتحنين نحو الصياغة الصحيحة، واقتراح التعليمات، وعرض التوضيحات والتعديلات على الفقرات أثناء فترة تطبيق الإختبار.

2.I : فقرات المطابقة (المزاوجة) Matching Items

تتكون فقرة ال مزاوجة من قائمتين. الأولى تسمى المقدمات premises ، والثانية تعرف بالاستجابات responses. بحيث يكون لكل مقدمة من المقدمات إجابة واحدة من القائمة الثانية. ويعبر عن القائمة الأولى بأرقام، بينما يعبر عن الإجابات بحروف أو العكس.

ويمكن أن يعرف هذا النوع من الفقرات بأسئلة التصنيف. لأنه يشيع استخدامها في الأسئلة التي تتطلب قدرة في التصنيف مثل (دول وعواصم) و (كتب ومؤلفون) و (أنهار وقارات) و (أحداث وتواريخ) و (رموز وعناصر) . كما يشيع استخدامها في حالة قياس فهم عناصر الخريطة وتحديد المواقع عليها، وفي تحديد أجزاء الصور. وفيما يلي عدد من المرشدات الواجب اتباعها عند صياغة فقرات المزاوجة:

*. أن تكون عناصر العمود الواحد متجانسة، لأنه عندما يقل التجانس بين المفاهيم، تصبح الفقرة مكشوفة وعديمة الأهمية.

*. أن تكون العناصر في العمودين قصيرة ما أمكن ، وإذا كان لا بد من العبارات الطويلة، عندها يفضل أن تكون الفقرات من نوع الإجابة القصيرة أو التكميل.

*. أن يكون عدد العناصر في عمود الإجابات أكثر من عدد العناصر في عمود المقدمات، كي لا يكون التطابق أو الإسقاط في النهاية وسيلة لمعرفة آخر إجابة.

*. أن يتراوح عدد العناصر في العمود الواحد بين (5) وحدات في عمـود المقدمات كحد أدنى، و(10) وحدات كحد أعلى في عمود الإجابـات. أما إذا زاد عن هذا الحد، فيفضل بناء سؤال آخر.

*. أن يكون ترتيب العناصر في عمود الإجابات عشوائيا، حتى لا يضيع الكثير من وقت الممتحن في استنتاج العلاقات بين ترتيب العناصر في عمود الإجابات والمقدمات.

*. أن يتم صياغة الفقرة كاملة في صفحة واحدة، وإذا كان السؤال طويلا، فيمكن تقسيمه إلى سؤالين أو أكثر.

مثال: وفق بين الدولة وعاصمتها في الجدول التالي:

الدولة	العاصمة
( ) هولندا	أ. أوسلو
( ) النرويج	ب. زيورخ
( ) الدانمارك	ت. أمستردام
( ) فنلندا	ث. هلسنكي
( ) السويد	ج. بروكسل
( ) بلجيكا	ح. برلين
	خ. استوكهولم
	د. كوبنهاجن

I.3 : أشكال أخرى من فقرات الإختيار من متعدد :

1- فقرات الاختيار من متعدد المركبة (Thorndike and Hagen, 1986)

بين هيوز وترمبل (Hughers and Trimble, 1963) أن هذا النوع من الفقرات أكثر صعوبة وأكثر قدرة على التمييز من فقرة الاختيار من متعدد العادية. ويشيع استخدام هذا النوع من الفقرات في تقويم المعرفة أو التطبيق أو التعليل في مجالات تشمل تعدد السببية، والوظائف والاستعمالات.

مثال: تفحص كلا من العمليات التالية في تصنيع الطعام أو إنتاجه:

1. سيدة أسرة تصنع الحلويات وتبيعها في بلدتها

جزار يصنع اللحوم المقددة والنقانق ويبيعها في ولاية مريلاند

مصنع للأطعمة يصنع الحلويات في ولاية كاليفورنيا ويبيعها في ولاية نيويورك

مسلخ للحوم في شيكاغو يشحن اللحوم إلى جميع الولايات

مزارع ينتج البرتقال في ولاية فلوريدا ويشحنه إلى الولايات الشرقية

أي العمليات المذكورة يخضع لتشريعات ورقابة الإدارة الفدرالية للأطعمة والأدوية؟

أ) 2 و 4 فقط

ب) 3 و 5 فقط

ج) 3 و 4 و 5 فقط

د) 1 و 2 و 5 فقط

هـ) جميعها

2- فقرات العبارات المتزاوجة كمثيرات (Thorndike and Hagen, 1986)

يشكل هذا النوع من الفقرات طريقة فعالة وجيدة في قياس القـدرة علـى الحكم على التفاوت بين كميات، أو على التأثيرات الناتجة عن التغير فـي أحوال وظروف، أو على التسلسل الزمني لوقائع. وعلى كل حال، فإن هذا النوع من الفقرات يفضل استخدامه للمضامين التي يكون لها قيم كمية ثابتة. والمثال التالي يوضح ذلك:

مثال على الفقرات ذات العبارات المتزاوجة كمثيرات

تعليمات الإجابة: تعتمد العبارات 1، 2 ، 3 على عبارات متزاوجة. اقرأ كـلا منها بعناية ثم أشر على ورقة الإجابة على:

أ – إذا كانت الكمية في العبارة الأولى أكبر مما هي في العبارة الثانية

ب– إذا كانت الكمية في العبارة الثانية أكبر مما هي في العبارة الأولى

ج– إذا كانت الكمية في العبارة الأولى مساوية للكمية في العبارة الثانية

1- الأولى: القيمة الحرارية في ملعقة كبيرة من سكر القصب.

الثانية: القيمة الحرارية في ملعقة كبيرة من الزبدة.

الجواب هو (ب)

2- الأولى: ما تحتاجه فتاة عمرها 16 سنة من الحديد يوميا.

الثانية: ما يحتاجه فتى عمره 16 سنة من الحديد يوميا.

الجواب هو (أ)

3- الأولى: السعرات الحرارية التي يحتاجها رجل يعمل في البناء وطولـه 185 سم ووزنه 85 كيلوغرام وعمره 25 سنة.

الثانية: السعرات الحرارية التي يحتاجها رجل يعمـل فـي الاسـتقبال وطوله 185 سم، ووزنه 85 كيلوغرام وعمره 25 سنة.

الجواب هو (أ)

II. الفقرات الصياغية

هي الفقرات التي يتم الإجابة عليها بأن يقوم الممتحن بصياغة وكتابة الإجابة بنفسه. وتعرف أحيانا بالفقرات الإنشائية أو المقالية. وهي التي لا تزال الأكثر شيوعا في تقييم أداء الطلبة في المدارس والجامعات. وهناك عدد من أنواع الفقرات ممن يندرج تحت هذا النوع، مثل فقرات التكميل والإجابة القصيرة، والفقرات الإنشائية وفقرات حل المسألة.

II. 1: فقرات التكميل والإجابة القصيرة Completion and Short Answer Items

وفيها يطلب من المفحوص أن يعطي جوابا مختصرا قصيرا يتلاءم مع السؤال المطروح، أو قد يكون المطلوب تعبئة فراغ في جملة أو نص معين. وهناك عدد من التوجيهات والإرشادات الواجب أخذها بعين الاعتبار عند صياغتها كما يلي:

*. **موقع الفراغ المطلوب تعبئته**. ويعتمد موقع الفراغ على كون الفقرة سؤالا أو جملة خبرية. فقد يكون الفراغ في النهاية إذا كانت الفقرة سؤالا وقد يظهر في البداية أو في مواقع مختلفة إذا كانت العبارة جملة. وبشكل عام ينصح أن يكون الفراغ في نهاية الفقرة، حيث أن وجوده في أول العبارة يزيد في تعقيد وغموض الفقرة، كما يستوجب ذلك من الممتحن أن يعيد قراءة الفقرة ليتمكن من الإجابة عليها كما في المثال التالي:

صياغة ضعيفة	صياغة مقترحة
يعرف ------ بأنه المدى الذي يستطيع إليه الإختبار أن يحقق الأهداف التي وضع لقياسها	تعرف قدرة الإختبار على قياس ما وضع لقياسه بـ ....الإختبار.

*. **عدد الفراغات:** يفضل صياغة عدد من الفقرات التي يحوي كل منها فراغا واحدا بدلا من المبالغة في عدد تلك الفراغات. إذ أن تعدد الفراغات في الفقرة الواحدة يجعل إجابتها أكثر صعوبة ويزيد في غموضا وينتفي البعد التربوي والتعلمي جراء الفقرة.

صياغة ضعيفة	صياغة مقترحة
في البحث ----- يقوم الباحث في التحكم بالمتغير المستقل، ويعتمد أسلوب ------ في عملية تكوين عينة الدراسة وفي تصنيفهم في مجموعات ----- ومجموعات -----.	في البحث التجريبي يتحكم الباحث أو المجرب بالمتغير-----، ويعتمد أسلوب العشوائية في عملية تكوين عينة الأفراد وفي تصنيفهم في مجموعات تجريبية و-----.

*. **طول الفراغ:** إن وجود تباين في طول الفراغات في الفقرة الواحدة أو في فقرات الإختبار يجعل الإجابة عليها معتمدا على التقدير والمقارنة بين الفراغات المطلوبة والعبارات التي يتذكرها الممتحن. الأمر الذي يجعل الممتحن يقوم بتسكين الإجابات في الفراغات التي تتناسب مع طولها، فيضع الجملة الطويلة في المكان الطويل، ويكتب الكلمة الصغيرة في فراغ يتلاءم مع طولها، وهذا ضرب من الإجابة بالتخمين. لذلك فإن جعل الفراغات بأطوال متساوية في كل فقرة أو حتى في جميع الفقرات يعد أمرا مفضلا كما في المثال التالي:

صياغة ضعيفة	صياغة مقترحة
---هو أكبر القيم تكرارا في توزيع البيانات ، بينما يعرف------------ --------- بأنه الجذر التربيعي لمتوسط مربعات انحرافات القيم عن وسطها.	--- هو أكبر القيم تكرارا في توزيع البيانات ، بينما يعرف ---بأنه الجذر التربيعي لمتوسط مربعات انحرافات القيم عن وسطها.

فإذا كانت المفاهيم التي ينوي الطالب ملأها في الفراغين هي (الانحراف المعياري) و(المنوال) فإن الممتحن في الصياغة الضعيفة سيختار المنوال للفراغ الأول والانحراف المعياري للفراغ الثاني لتناسب طول الفراغ مع حجم المفهوم.

***. الفقرات الإنشائية**

تعرف الفقرات الإنشائية **Essay Items** أحيانا بالفقرات المقالية كما ذكر سابقا. وقد بدأ استخدام هذا النوع من الفقرات منذ فترة غير قصيرة، أو منذ بدأ التقييم. وبلغة أخرى، فقد بدأ العمل به قبيل الفقرات الموضوعية، ومازال الكثير من التربويين والأكاديميين يعتبر السؤال المقالي هو الأصل في عملية تقييم أداء الطلبة والممتحنين. ومهما يكن من أمر فإن هناك عدد من الإرشادات التي يجب الاطلاع عليها عند البدء بكتابة فقرات الإختبار الإنشائية كما يلي:

***. أن يكون المطلوب في السؤال محددا بوضوح:** بمعنى أن لا تكون الفقرة مفتوحة بشكل مطلق. وبغير هذا يشعر الممتحن بأنه غير مقيد ولن يتوقف في إجابته عند نقطة معينة. الأمر الذي يصبح تصحيح مثل هذا النوع من الفقرات أمرا صعبا للغاية. ولهذا ينصح بتحديد إطار مفصل للإجابة، كالقول مثلا:

- أكتب أفضل تعريف لصدق أداة القياس؟
- ما الفرق بين المناعة الإيجابية والمناعة السلبية؟
- قارن بين الإختبارات محكية المرجع ومعيارية المرجع من حيث تفسير الأداء.
- ما الأسباب الاقتصادية التي أدت إلى نشوب الحرب العالمية الثانية؟

أما إذا تعذر كتابة سؤال مقالي محدد يسهل قياسه، فلا بد من تقسيم السؤال الواحد إلى عدد من الأسئلة المستقلة عن بعضها البعض. أو أن يتم تحديــد السؤال الكبير بشروط الإجابة من مثل:

- قارن بين العلاقة الخطية وغير الخطية من خلال التمثيل البياني لكل نوع منها.
- صنف الفقرات التالية في ضوء قيم معاملات الصعوبة.
- أي المتغيرات التالية هو الأكثر تنبؤا بمستوى أداء العاملين فــي المصارف التجارية ؟

***. تجنب أن يحتوي الإختبار على أسئلة اختيارية**

أي أن يطلب من الممتحن أن يجيب على عدد من الأسئلة من بــين عدد أكبر (ثلاثة أسئلة من أصل خمسة مثلا). وهذا ربما يؤدي إلى أن يهتم الممتحنون ببعض أجزاء المادة الدراسية ويهملون البعض الآخر. كمــا أن هذا النوع من الأسئلة ربما يضعف عملية المقارنة بين الممتحنــين لعــدم تساوي أسئلة الامتحان كافة بمستوى الصعوبة أو الأهمية. إضافة إلى أنها قد تربك بعض الممتحنين في عملية الاختيار.

***. ابدأ عبارة سؤال المقال بأفعال قابلة للقياس من مثل:**

(وضح كيف....)، (ميز بين.....)، (أعط أمثلة من عندك....)، (بــين الفرق .....)، (أنقد...)، (بين مــا ذا تتوقــع أن يحــدث إذا .....)، (قارن بين .....).

***. المفروض أن يأخذ واضع الإختبار وقتا كافيا لإعداد الفقرات الإنشائية.**

لذلك من المهم أن يتم تحليل المادة العلمية أو المجالات السلوكية التي سيغطيها الإختبار إلى عناصر وأقسام فرعية، ثم يعمل على كتابة

أسئلة تغطي كافة هذه الأقسام بدرجات ونسب متفاوتة. وهذا بلا شك يتطلب وقتا كبيرا.

*. لا ينصح باستخدام الفقرات الانشائية إذا كان من الممكن استخدام الفقرات الموضوعية أو فقرات الإجابة القصيرة. وذلك لتمتع الأنواع الأخيرة بدرجة أفضل من الثبات وقدرتها على تغطية مجالات سلوكية أوسع.

بناء الإختبارات محكية المرجع

قد يعجب المعلمون بالإختبارات محكية المرجع لأنها طيعة ومن الممكن أن تتماشى مع ما يقومون بتدريسه وما يتوقون لاختباره. كما أنها توفر معلومات تشخيصية مهمة للمعلم والطالب على حد سواء، تفيد في تطوير خطط مستقبلية لتطوير البرامج التربوية.

لا تختلف إجراءات وطرائق ومبادئ كتابة فقرات الإختبارات محكية المرجع كثيرا عن تلك التي يجب مراعاتها عند كتابة فقرات الإختبارات معيارية المرجع. إذ لا يستطيع الفرد التمييز بين فقرة محكية المرجع وأخرى معيارية المرجع بمجرد النظر إليها.

وفي كلا الحالتين، ينبغي كتابة فقرات الإختبارات في ضوء جدول مواصفات معين، كما يجب أن يعنى كلاهما بالصدق والثبات، ويفترض أن يساعدا المستخدمين على اتخاذ قرارات معينة بشأن الأفراد مع اختلاف في السياق الذي تتخذ فيه هذه القرارات. وبصورة عامة، لا نتوقع أن تكون للاختبارات محكية المرجع أية معايير، مع أنه ليس غريبا أن تمتلك هذه الإختبارات بعض المعلومات المعيارية **(Popham, 1987)**.

أما الاختلاف الرئيسي ما بين النوعين فيتلخص في أن بؤرة اهتمام الإختبار محكي المرجع أضيق من بؤرة اهتمام الإختبار معياري المرجع

(عدد أقل من الأهداف ولكنها تقاس بعمق ودقة من خلال كثـرة الفقـرات المتعلقة بكل هدف أو مهارة).

وعند بناء الإختبار محكي المرجع، لا يهتم الفرد بأن يكون هناك انتشار واسع (تشتت) لدرجات الطلاب. فبعد أن يتم تدريس الطلبة نأمل أن يقـوم جميعهم بالإجابة على كل الفقرات بشكل صحيح. كما يتوقع من الطالب في هذا النوع من الإختبارات أن يقوم بأداء عمل معين **perform**، كأن يقـوم مثلا بموازنة معادلة كيميائية، أو بحل معادلة تفاضلية، أو القيام بحل مسألة تتطلب إجراء القسمة الطويلة. وقد اقترح أن تكون فقرات هذا النوع مـن الإختبارات من نوع الإجابة القصيرة أو المقال. في حين تـستخدم معظـم الإختبارات معيارية المرجع الفقرات الانتقائية كفقرات الصواب والخطأ، أو الاختيار من متعدد (**Hills, 1981**). وهناك عنصران آخـران لا بـد مـن أخذهما بالحسبان عند بناء الإختبارات محكية المرجع هما:

(1) عدد الفقرات اللازمة لقياس كل هدف أو مخرج. فإذا كانت هنـاك ثمة قرارات يجب اتخاذها، فإن يمكن الاكتفاء بعشر إلى عـشرين فقرة لكل هدف أو مخرج تعليمي يراد قياسه.

(2) المحك أو الدرجة الفاصلة **cut-score** التي تدل على مـستوى الإتقـان. وهناك عدد من الطرق والأساليب التي تستخدم لاشتقاق تلـك الدرجـة حسب عدد من العوامل.

بناء اختبار محكي المرجع (تشخيصي)

تتنوع أشكال التقييم التي تستخدم في إطار العملية التعليميـة التعلميـة وتختلف في ضوء الأهداف والنتائج المنوي التوصل إليها. ولعـل التقيـيم محكي المرجع هو أكثر تلك الأشكال مساهمة في تخطيط التعليم على نحو فاعل. إذ يهتم هذا النوع من التقييم في:

- قياس وتشخيص ما حققه المتعلم من مهارات ومعارف، وما اكتسبه من اتجاهات وسلوكات تتعلق بمجال سلوكي تدريبي أو تعلمي في ضوء محك أو مستوى معين.
- الكشف عن الصعوبات التعلمية ومواطن الضعف التي لم يتم التغلب عليها باستخدام التغذية الراجعة التي تم تطويرها أثناء عملية التقويم البنائي.
- تقدير علامة المتعلم في مجال محدد. وبشكل مبسط، يمكن القول بأن نسبة الفقرات التي أجاب المفحوص عنها إجابة صحيحة هي علامته على ذلك المجال.
- تصنيف المتعلمين إلى فئات متمايزة حسب مستوى الإتقان **Mastery Level**. وقد يكتفى أحيانا بالتصنيف في:
  - فئتين متمايزتين هما: **متقن Master** وهو من كانت علامته تساوي أو أعلى من درجة القطع، **وغير متقن Non-Master** إذا حصل على علامة دون درجة القطع.
  - أكثر من فئتين متمايزتين. وعليه يجب تقسيم مقياس علامة المجال موضع البحث إلى فئات إتقان متمايزة عددها **K** فئة. تحدد بنقاط قطع عددها **(k-1)**.

وجدير بالذكر القول بأن التصنيف في فئتين متمايزتين هو الأكثر شيوعا في حالة اختبارات التشخيص.

خطوات بناء الإختبار التحصيلي محكي المرجع

1. تطوير مواصفات تفصيلية شاملة للاختبار ويتضمن ذلك المراحل التاليه:

a. **التخطيط للاختبار ويتطلب ----**

i. تحديد الغرض من الإختبار: **تشخيص مواطن ضعف وقوة التلاميذ ومن ثم التحقق من أسبابها.**

ii. تحديد مجال المحتوى الذي سيقيسه الإختبار: **(الكسور مثلا).**

iii. تحديد الفئة المستهدفة بالإختبار. من حيث الفئة العمرية، الصف الدراسي، الشعبة أو الشعب الدراسية **(طلبة الصف الخامس الابتدائي).**

iv. تحديد أعضاء فريق العمل وتوزيع المهام على كل منهم.

b. **تحليل المحتوى الذي يعمل عليه الإختبار:**

i. تحليل موضوع قسمة الكسور تحليلا تفصيليا وتحديد عناصر المعرفة المتعلقة به، انطلاقا من بداية المرحلة الدراسية التي قرر فيها موضوع الإختبار **(المرحلة الابتدائية).**

ii. تنظيم عناصر المعرفة المتعلقة بموضوع الإختبار **(الكسور)**

1. تحديد المتطلبات السابقة من عناصر المعرفة الرياضية اللازم توافرها لدى المتعلم كي يحقق مهارة "**قسمة الكسور العادية**"

2. ترتيب المهارات الفرعية ترتيبا منطقيا بحيث يكون إتقان المتعلم لواحدة أو مجموعة منها متطلبا أساسا لإتقان مهارة أخرى أعلى منها مرتبة.

3. وضع التحليل على هيئة شجرة توضح البنية الهرمية للموضوع.

c. **التحقق من صدق التحليل وصدق البناء الهرمي ويتضمن:**

i. التحقق من صدق البناء الهرمي لعناصر الموضوع عن طريق:

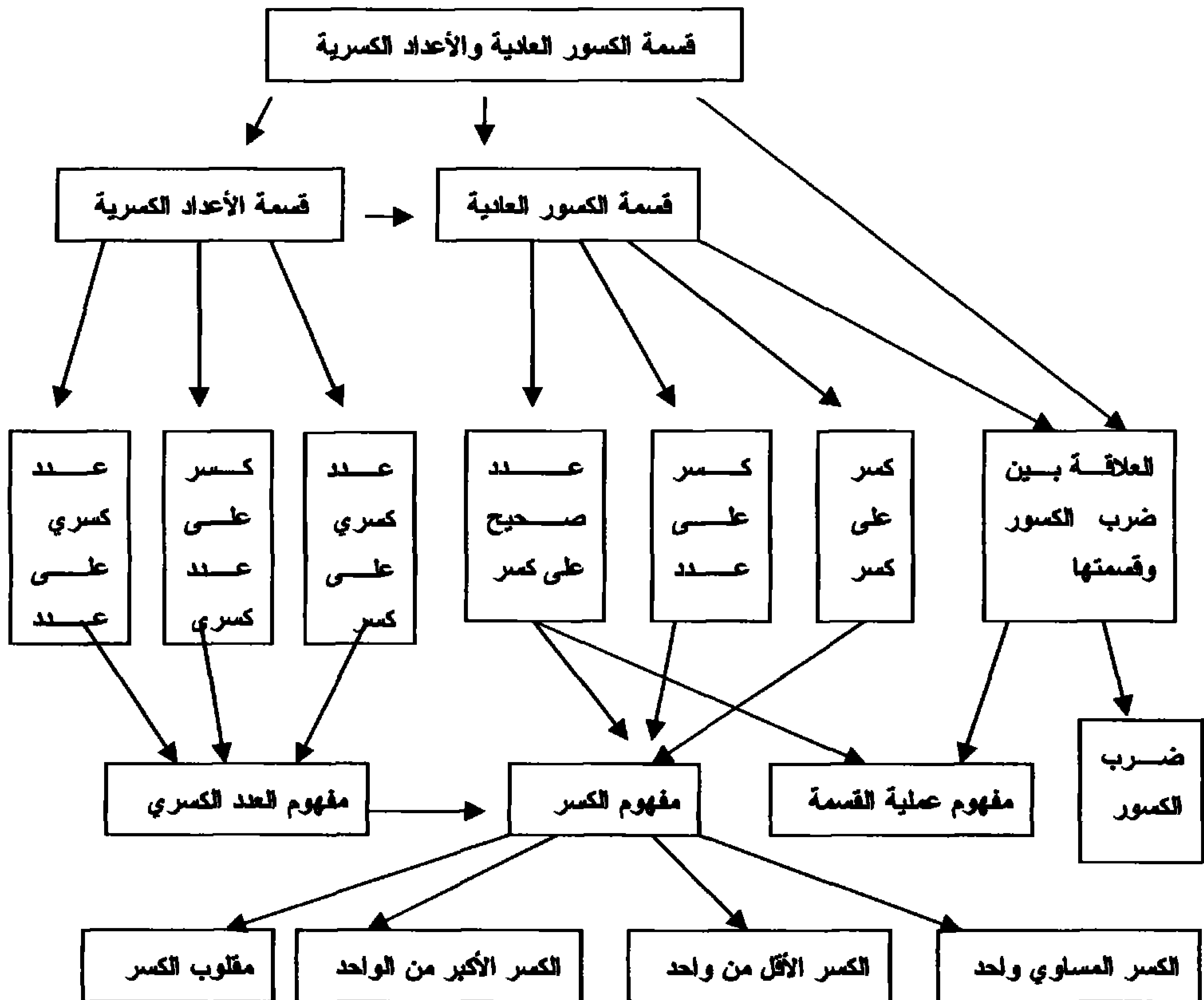
1. عقد لقاءات متعددة مع فريق العمل.
2. عرض البنية الهرمية المتعلقة بموضوع الإختبار على فريق العمل والطلب منهم الحكم عليها من حيث:
   - صحة المعلومات الواردة فيها.
   - انتمائها إلى الموضوع.
   - دقة التسلسل الهرمي للعناصر.
   - صحة العلاقات التي تربط بين عناصر البناء الهرمي.
3. تفريغ النتائج ومن ثم تحليلها بعد تحويل التقديرات إلى تقديرات ثنائية. ومن ثم حساب درجة الارتباط بين تقديرات كل اثنين من المحكمين.

ii. التحقق من صحة الفرضيات المتعلقة بمستويات التحليل في البناء الهرمي.

نظرة إضافية نحو تخطيط الإختبار

بعد أن يتم تحديد الهدف من الإختبار، يجري العمل نحو تحديد شكل الفقرات التي سوف يتم تطويرها. هنا ينبغي استعراض عدد من البنود الأساسية التالية المتعلقة بفقرات الإختبار وإخراجه:

- طول الإختبار أو عدد فقراته. ويتحدد هذا في ضوء التالية:
  - الغاية والغرض من الإختبار
  - أنواع الفقرات المستخدمة
  - درجة الثبات المطلوبة
  - الفئة العمرية للمستهدفين بالإختبار



- مستوى قدرة المفحوصين
- الوقت المخصص للاختبار
- طبيعة الجهد المطلوب للإجابة على الفقرة
- الأهداف التعلمية المراد تحقيقها

- مستوى الصعوبة التي يتصف بها الإختبار
- توقيت وعدد مرات إجراء الإختبار
- طبيعة المرفقات (المعينات): صورية، كلامية، أم غيرها
- الامتحان بكتاب مفتوح أو بكتاب مغلق

طول الإختبار

ليس هناك من معادلة **formula** ترشد مطور المقياس أو الإختبار (المعلم مثلا) عن العدد المثالي للفقرات والعبارات. ولكن، وبشكل عام، من المفيد القول بان العدد الإجمالي للفقرات يجب أن يكون كافيا لتوفير تمثيل شامل لمجالات السمة أو المحتوى، ويغطي عددا معقولا من الأهداف أو المخرجات المنوي تحقيقها. وبالرغم من نية المعلم، "وخصوصا في الإختبارات محكية المرجع" إعطاء كل مفحوص الوقت الكافي لبيان معرفته بالموضوع، يجب أن يكون هناك زمنا محددا للاختبارات الصفية لأسباب عملية. ويختلف طول وقت الإختبار في ضوء النقاط التالية:

الغاية والغرض من الإختبار: فإذا كان الإختبار يغطي محتوى تعليمي يتوزع على فصل دراسي، فان ذلك يتطلب فقرات أقل. أما عندما يكون اهتمام المعلم الوصول إلى تغطية شاملة ومكثفة للمادة الدراسية في اختبار تشخيصي، يكون لزاما عليه استخدام اختبار أكثر شمولية. كما أن مسالة استخدام الإختبار لأغراض التقييم البنائي أو التقييم التجميعي النهائي، تلعب دورا كذلك في تحديد طول الإختبار.

أنواع الفقرات المستخدمة: يحتاج السؤال المقالي إلى وقت اكثر مما تحتاجه الفقرة الموضوعية وتحتاج فقرات الإجابة القصيرة إلى وقت أطول مما تحتاجه فقرات الصواب والخطأ. أما فقرات الصواب والخطأ فتتطلب وقتا أقصر نسبيا من فقرات الاختيار من متعدد. أما الخطوط الدلالية والإرشادية التي نستطيع توفيرها في ما يخص طول الفقرة بالعلاقة مع نوعها وشكلها، فهي كما يأتي:

في فقرات الاختيار من متعدد ذات الأربع أو الخمس بدائل والمستخدمة مع الصفوف في المرحلة الثانوية، فيجب أن يكون غالبية الطلبة قادرين على إجابة الفقرة في (75) ثانية. فإذا كان زمن الإختبار (50) دقيقة مثلا، يستطيع المعلم أن يخطط لاستخدام (35) فقرة اختيار من متعدد ذات خمس بدائل. وعلى الرغم من أن حوالي (45) دقيقة فقط هي ما يستخدم لهذا الإختبار، إلا أن عددا من العوامل ربما تؤثر على ذلك من مثل درجة تعقيد المادة الدراسية أو الهدف التعليمي المراد قياسه.

في الإختبارات المقالية القصيرة (حوالي نصف صفحة للسؤال الواحد)، يستطيع معظم الطلبة الإجابة على حوالي ستة أسئلة خلال ساعة واحدة تقريبا. وتتنوع هذه التوقعات تبعا لطبيعة الأسئلة ومجالات المحتوى وعمر المفحوص.

أما في الإختبارات المقالية الطويلة (صفحتين أو ثلاثة للسؤال الواحد)، فيستطيع معظم طلبة الصفوف الأولى أو المتقدمة من المرحلة الثانوية الإجابة على حوالي ثلاث أسئلة في الساعة الواحدة. بينما يستطيع الفرد في الصفوف الرابعة والخامسة، أن يعطى سؤالين مقاليين طويلين في فترة

اختبار مكونة من ساعة واحدة تقريبا. وإذا شعر المعلم بضرورة استخدام اختبار مقالي أطول، وكانت أعمار الطلبة وقدرتهم تستوجب اختبارا قصيرا، يستطيع المعلم آنذاك تقسيم الإختبار وإجراءه في أيام متتالية. وبشكل عام، يحتاج المفحوص (50) ثانية تقريبا للإجابة على الفقرة الواحدة من نوع الإجابة القصيرة أو المطابقة أو الصواب والخطأ.

وجدير بالذكر، بأن يأخذ واضع الإختبار التوقيتات السابقة الذكر على أنها مقترحات فقط، ففي بدء الفصل الدراسي أو في وقت مبكر منه، بوسع المعلم أن يجري اختبارا لطلبته ويسجل الوقت اللازم للغالبية (80 %) منهم لإكمال الإختبار. ويلزم نفسه وفقا لذلك في الإختبارات المستقبلية اللاحقة مع هؤلاء الطلبة.

درجة الثبات المطلوبة: سنرى لاحقا بأنه كلما كان الإختبار طويلا، مال نحو الثبات. وطالما أن الأخطاء العشوائية يحذف بعضها البعض الآخر، فإن درجة الثبات تتأثر بطول الإختبار. وبالرغم من أن كثرة الفقرات قد تزيد من ثبات الإختبار، إلا أن علينا أن ندرك أن التعب والإجهاد قد يؤثران سلبا على درجة الطالب بشكل ملحوظ.

الفئة العمرية للمستهدفين بالإختبار: يميل الأطفال الصغار إلى القراءة والكتابة وربما التفكير أيضا بشكل أبطأ من الكبار. كما يميل الأطفال إلى الحركة وعدم الاستقرار بمعدل أكبر من الكبار. وعليه لا يمكن جعل اختبارات تلاميذ المرحلة الابتدائية بنفس طول اختبارات المراحل المتوسطة والثانوية. ومن القواعد العامة التي أكدتها التجربة أن نحاول تحديد زمن الإختبار بثلاثين دقيقة لتلاميذ المرحلة الابتدائية. وحوالي (40) دقيقة لطلبة الصفوف الأولى من المرحلة المتوسطة. ويتوقع أن يصل زمن الإختبار لطلبة الصفوف المنتهية من المرحلة الإعدادية على حوالي

(90) دقيقة. كما يمكن أن يكون ذلك ساعتان أو أكثر لطلبة المرحلة الثانوية وطلاب الجامعة.

مستوى قدرة المفحوصين: مثلما يحتاج الأطفال الصغار إلى وقت للاستجابة لفقرة الإختبار أطول من الوقت الذي يحتاجه الكبار، يحتاج الأطفال بطيئو التعلم أيضا إلى وقت أطول من الذي يحتاجه التلاميذ الاعتياديين أو الموهوبين. يجب أن يعرف المعلم تلاميذه ويكون ملتزما ليس فقط بشكل الفقرة المستخدمة، بل وبخصائص طلبته فيما يتعلق بطول الإختبار.

الوقت المخصص للاختبار: يجب أن تكون أغلب الإختبارات التحصيلية، وخاصة المعرفية (وليست المهارية) اختبارات قوة أكثر منها اختبارات سرعة. ووفقا لذلك، يكون من المتوقع من كافة المفحوصين أن يكونوا قادرين على المحاولة مع فقرات الإختبار في الوقت المسموح به. وبالرغم من أن السرعة في الإجابة تعد معيارا يقيم الأداء، إلا أنه في أي اختبار تحصيلي نحاول قدر الإمكان معرفة درجة أو مقدار معرفة الطلبة وليس سرعته في الإجابة.

طول وتعقيد الفقرة: كلما كثرت المواد المرافقة أو المعينة (مثل الخرائط أو الجداول أو تقارير التجارب العلمية)، كلما تزايد الوقت المطلوب للقراءة. وعليه، يقل عدد الفقرات التي يمكن طرحها في الإختبار في الوقت المحدد.

طبيعة الجهد المطلوب للإجابة على الفقرة: فقد تحتاج بعض فقرات الإختبار إلى الكثير من الحساب الرياضي والمعالجات الكمية. عندها، يجب التعامل مع عدد أقل من الفقرات التي تعطى في وقت محدد على عكس

الفقرات الشفهية. هذه الحقيقة تشابه إلى حد كبير ما تم مناقشته عن نوع وشكل الفقرة وتأثيرها على وقت الإختبار.

الأهداف التعليمية المراد تحقيقها: إن الفقرات التي تتطلب تذكر المواد والحقائق فقط، يمكن الإجابة عليها بسرعة أكثر من تلك الفقرات التي تحتاج إلى فهم أو مهارة فكرية عالية المستوى. وبشكل عام، فإن قياس الوقت الكافي لإكمال الإختبار أمر يتطور بالتجربة والمران، ويعتمد كثيرا على طبيعة المحتوى وعلى مهارة المعلم في كتابة فقرات واضحة وموجزة وغير غامضة. ويجب على كل معلم من خلال التجربة أن يحدد حدود الفترة الزمنية العملية والمناسبة للتلاميذ وللاختبار.

مستوى الصعوبة التي يجب أن يتصف بها الإختبار: يستطيع المعلمون جعل اختباراتهم سهلة جدا، أو صعبة جدا، أو متوسطة الصعوبة. ويشعر بعض المعلمين بأنهم سيكسبون احترام تلاميذهم عن طريق إخضاعهم لاختبارات سهلة. كما يشعر نفر آخر من المعلمين بأنه كلما كان الإختبار صعبا، كان ذلك أفضل، من منطلق أن الإختبار الصعب هو الذي يكسبهم احترام الطلبة وأولياء أمورهم على السواء. إن كلا المجموعتين من العلمين على خطأ. عدا عن أن الشيء الإيجابي الوحيد الذي نعرفه عن الإختبارات الصعبة هو أنها تجعل الطلبة يهتمون أكثر. ويعتمد مفهوم الصعوبة أو ما يحدد صعوبة الإختبار على عدد من العوامل، وأبرزها:

- **الغرض من الإختبار**
- **مستوى قدرة الطلبة**
- **الفئة العمرية للطلبة المستهدفين**

وهذه هي نفس العوامل الواجب مراعاتها عند التخطيط لعدد الفقرات وشكل الفقرة المستخدمة في الإختبار. ونحن لا نعتقد أن هناك ضررا من

إعطاء الطلبة اختبارا سهلا، خصوصا إن كان الهدف الوحيد منــه تنميــة وتعزيز مشاعر الثقة بالنفس واحترام الذات لدى بطيئي التعلم من الطلبة.

وهناك حيثيات توجب استخدام الإختبارات الصعبة. حيث أنــه مــن المفيد جدا للطلبة الجيدين من أجل إعــدادهم لاختبــارات صــعبة لاحقــة كاختبارات القبول والانتقاء لدخول كلية أو تخصص مميز، شــريطة أن لا تكون الإختبارات غاية في الصعوبة إلى الحد الذي قد يرعب الطلبة ويرفع القلق عندهم.

وينطوي مفهوم الصعوبة على الكثير من المعاني فــي الإختبــارات الموضوعية اكثر مما يعنيه في الإختبــارات المقاليــة أو الــشفهية. ففــي الإختبارات الموضوعية، يكون الجواب إما صحيحا أو خاطئا. أما في البقية فيمكن أن يحدث تفاوت في درجات الصحة، كما أن لصعوبة الفقرة علاقة بما نهدف من الإختبار أن يميز بين الطلبة وفقا لمستوى تحصيلهم اكثر مما لو أردنا منه أن يكون اختبارا تشخيصيا **Diagnostic** أو اختبــار إتقــان **Mastery**. ففي الإختبار التشخيصي، لا نكون معنيــين بمقارنــة الفــروق النسبية بين طالب وآخر، بل نعنى بمعرفة مواطن قوة وضعف كل منهمــا من أجل أن نوصي بالتدريس العلاجي إن تطلب الأمر.

وفي اختبار الإتقان، نتوقع أن يكون الإختبار سهلا نسبيا أيضا، لأنه من المفترض بان المعلم قد قام بالتدريس بشكل جعل كل أو أغلب التلاميذ يتعلمون بعض الأساسيات (المهارات أو المعلومات) ويسيطرون عليها.

وإذا كان الإختبار قبليا (بمعنى أنه يتم تطبيقه قبل القيام بتدريس وحدة أو مفهوم معين في الصف) من اجل تطوير وبناء مواد تدريــسية فاعلــة، فنتوقع أن تكون درجات اغلب الطلبة ضعيفة، لأنهم لم يتعلموا تلك الوحدة أو ذلك المفهوم بعد. وعليه، لسنا معنيين في اختبارات الإتقان والإختبارات

التشخيصية والقبلية بالصعوبة، لأنه ليس غايتنا التمييز بين الطلبة. وفي حالات الانتقاء أو التصنيف، فإننا نحتاج إلى اختبار بمساحة واسعة من الدرجات تتيح ملاحظة التفاوت بين المفحوصين.

ومن أجل الحصول على المعلومات التي تساعد المعلم في عملية التمييز بين الطلبة حسب تحصيلهم النسبي، على المعلم أن يعد أسئلة نتوقع أن يجيب عليها حوالي نصف الطلبة بصورة خاطئة، إذا لم يحصل أي تخمين لديهم، (حينئذ نستطيع القول أن الإختبار ذو صعوبة متوسطة).

ويعتقد كذلك بأنه حتى الإختبار الذي يستخدم لأغراض تمييزية يجب أن يحتوي على بضع فقرات سهلة جدا وبضع فقرات صعبة جدا. وإذا ما تم اتباع هذه الخطة فان الطلبة الضعفاء سوف يتحفزون على الاستمرار، خصوصا إذا جاءت الفقرات السهلة في بداية الإختبار. بينما تتحدى الفقرات الصعبة قدرات الطلبة اللامعين، لكن أغلبية الفقرات يجب أن تكون على مستوى متوسط من الصعوبة.

من الطرق الفاعلة في الحصول على تقدير صعوبة الفقرة هي تجريب الفقرات لغرض الاستطلاع الأولي على مجموعة من الطلبة يشبهون الطلبة الذين صمم الإختبار من أجلهم. ويمكن استخدام أساليب تحليل الفقرة في ما بعد للحصول على معاملات الصعوبة والتمييز. ولأن الإختبار الاستطلاعي صعب على المعلم إعداده وتطبيقه، يمكن الاعتماد على الأحكام الذاتية. وعادة ما يوصى المعلمون بتصنيف فقراتهم إلى (صعبة، متوسطة، أو سهلة).

باختصار، يجب أن تعتمد صعوبة الإختبار إلى حد كبير على الغاية منه. فإذا كانت نتائج الإختبار تستخدم في وصف أو تشخيص قدرات الطلبة (محكي المرجع)، فلن تكون درجة صعوبة الفقرة عاملا رئيسيا في اختيار

فقرات الإختبار. من ناحية أخرى، إذا أريد استخدام نتائج الإختبار للتمييز بين الطلبة على أساس تحصيلهم، يكون هناك أهمية أكبر لصعوبة الفقرة.

توقيت وعدد مرات إجراء الإختبار

عادة ما يتساءل المعلمون: "هل يجب أن نجري اختبارا واحدا في الأسبوع؟ أو أكثر؟". يفضل بعض المعلمين إجراء اختبارات متعددة بأجزاء من الفصل الدراسي في مناسبات متكررة. في حين يفضل البعض الآخر من المعلمين إجراء عدد أقل من الإختبارات، ولكن بوحدات واسعة من الفصل الدراسي. أما الآن، فليس هناك من دليل يبين أن الإختبار الذي يعتمد على أجزاء صغيرة من الفصل الدراسي المقرر افضل من الإختبار الذي يمثل وحدة أكبر وأوسع من العمل. من ناحية أخرى، يفضل البعض الإختبارات المتكررة لأنها توفر قاعدة رصينة ودقيقة لجعل المعلمين والطلبة وأولياء أمورهم على دراية بواقع التلاميذ وتقدمهم (**Mehrens and Lehman, 1991**).

ومرة أخرى، فإن استخدامات نتائج الإختبار هي التي تحدد تكرار الإختبارات. ففي أغلب البرامج التعليمية المبنية على الأهداف والتي تنسجم مع أساليب التقويم البنائي، يعطى التلاميذ اختبارات قصيرة محكية المرجع، وبصورة متكررة لتزود المعلم بصورة واضحة عن أداء التلاميذ وتحدد له حاجتهم إلى تعليم إضافي أو علاجي.

و تؤثر طبيعة التلاميذ على تكرار الإختبارات ومرات إجرائها. حيث يقوم المعلمون الذين يتعاملون مع التلاميذ بطيئي التعلم والأطفال الذين يحتاجون إلى تعليم إضافي علاجي ومع التلاميذ اللامعين في ذات الوقت بإجراء الإختبارات في فترات أو فواصل زمنية متكررة. ونرى بأن على المعلم أن يحدد مرات تكرار الإختبار.

وبصورة عامة، فإن اختبار الطلبة مرتين على الأقل فـــي الفـــصل الدراسي الواحد، ربما يكون مفيدا، على اعتبار أن التقييم المعتمد على أكثر من اختبار أفضل فعالية من ذلك المعتمد على اختبار واحد (Mehrens and Lehman, 1991).

كما أننا نعرف جيدا بان التكرار المتزايد للاختبارات يؤثر سلبا على وقت البرنامج التعليمي ويتجاوز عليه. وعلى أية حال، لا نتفق مـــع نقـــاد الإختبارات من الذين يعتقدون بأن وقت الإختبار ينقص من وقت التدريس. وربما يكون هذا صحيحا، إذا كانت الفقرات المستخدمة في الإختبار ضعيفة جدا. أو أن المعلم لم يستخدم نتائج الإختبار لغرض الحصول على معلومات حول مستوى قوة وضعف التلاميذ. أما إذا كانت الإختبـــارات المـــستخدمة فاعلة، وأخذ المعلمون وقتا كافيا لتحليل طبيعة الاستجابات ومن ثم مناقشتها مع الطلبة، فإن النتائج ستكون مصدرا مهما للمعلومات لكل مـــن المعلـــم والمتعلم، وكذلك المؤسسة التعليمية وأولياء الأمور.

طبيعة المرفقات أو المعينات

تعتمد طبيعة المواد المرفقة مع فقرة الإختبار إلى حد كبيـــر علـــى طبيعة المحتوى المراد اختباره، وعلى أعمار الطلبة الخاضعين للاختبـــار. فمع الأطفال الصغار، يوصى باستخدام الكثير من الصور والقليل من المواد المنطوقة (اللفظية) (ما لم نكن نريد قياس القراءة أو الاستعداد للقراءة). كما يوصى باستخدام المفردات البسيطة الملائمة لعمر وقـــدرة الطالـــب، وإلا أصبح الإختبار مجرد امتحان للقراءة أو الاستعداد العام.

إن أسئلة مثل: "كيف يجب أن نقدم فقرات الإختبار؟"، "هل ينبغـــي طبع الفقرات على شكل كتيب صغير أم يجب تقديمها في فلم صـــوري أو شرائح صورية أو أي طريقة إيضاح سمعية بصرية؟"، تبـــدو ذات صـــلة

بالموضوع. وقد أظهرت الأبحاث والدراسات أن الممتحنين (الطلبة) يكونون أكثر سرعة في معدل استجاباتهم عندما يجري الإختبار بوسائل الإيضاح السمعية والبصرية بل وأكثر سرعة مما لو كانوا يجيبون على فقرات الأسئلة في الإختبار المكتوب في كتيب، ويستطيعون العمل بسرعة أكبر مع مراعاة الدقة.

امتحانات الكتب المفتوحة والكتب المغلقة

يتوق أغلب المعلمين إلى زيادة الفرص أمام طلبتهم لعمل أفضل ما بوسعهم في الإختبارات التحصيلية الصفية. وهنالك بعض الخلاف حول الطريقة الأمثل للحصول على ذلك. فهنالك من يؤكد على أن على الطالب أن يكون قادرا على استخدام الوسائل المساعدة كالملاحظات والنصوص والمصادر الأخرى أثناء الامتحان (أي أن يكون الامتحان بكتب مفتوحة). ويذهب المعلمون الذين يؤمنون بهذا النوع من الامتحانات إلى أنها تحد من الغش **أولا**، ولان الكتب المفتوحة، لا تنوب عن التحضير للامتحان لان الوقت اللازم للنظر في الملاحظات والمصادر لغرض الإجابة سيثني الطلبة عن الاعتماد كليا على هذه المصادر **ثانيا**. ولأن هذه الإختبارات تجعل الطالب يدرس من اجل استيعاب وتطبيق المعلومة وليس لمجرد الحفظ الأصم للحقائق **ثالثا**.

أما المعلمون الذين لا يتفقون مع هذا ويزعمون أن مساوئ هذا النوع من الإختبارات تفوق قيمتها، يؤكدون على عدم السماح للطلبة باستخدام أي نوع من الوسائل المساعدة. ومهما يكن من أمر، تشير الدلائل إلى أن:

- **الطلبة يفضلون الإمتحانات مفتوحة الكتب لأنها أقل إثارة للقلق،**

- **هناك اختلاف ضئيل جدا في الأداء بين الطلبة الذين يخضعون للاختبارات مفتوحة الكتب وأولئك الذين يؤدون اختبارات مغلقة الكتب،**
- **الإختبار مفتوح الكتاب يقيس القدرات العقلية العالية في التفكير وإصدار الأحكام أكثر مما يؤديه الإختبار مغلق الكتاب،**
- **الطلبة يستعدون لهذين النوعين من الإختبارات بأساليب مختلفة،**

يبدو أن قرار المعلم باستخدام اختبار مفتوح الكتب أو مغلقه يعتمد بشكل كبير على أهدافه التعليمية. إذا كانت أهداف المعلم هي التحقق من مدى سيطرة طلبته على بعض الحقائق والمواد المعرفية (وهناك الكثير من الأمثلة التي يكون فيها حفظ الطلبة لبعض التفاصيل والحقائق أمرا لا مناص منه). يكون من المناسب استخدام امتحانات بكتب مغلقة. من ناحية أخرى، إذا كانت الأهداف تنصب على تقويم فهم الطلبة واستيعابهم وتطبيقهم لبعض المفاهيم، وتقويم قدراتهم على استرجاع المعلومات، وتقويم قدراتهم على كتابة تقرير مقنع، فهنا يكون أيا من نوعي الامتحان مناسبا.

التصحيح Scoring

إن معرفة الطالب المسبقة بكيفية تصحيح التقييم تساعده كثيرا في الإعداد، وخاصة عند الإجابة على المهام مفتوحة النهاية. فإذا اتجهت النية إلى خصم نقاط من الدرجة بسبب التهجئة أو الضعف الإملائي في كتابة بعض الأسماء والمصطلحات، فإن من حق الطالب أن يتدرب على كتابتها مسبقا وبشكل جيد.

كذلك، يحتاج الطلب أن يعرف كيف ومتى يمكنك منحهم درجات في حالة الإجابة غير الكاملة على بعض المهام، وما مقدار الوزن المخصص لكل مهمة. ولا يخفى أن إشراك الطلبة بتصميم دليل تصحيح

بشكل مسبق يعد عملا محمودا. وسيتم معالجة تصحيح أدوات التقييم بنوع من التفصيل خلال الفصول القادمة من هذا الكتاب.

استخدام النتائج

يحتاج أن يعرف الطلبة أهمية الدرجة على التقييم في القرار المنوي اتخاذه بشأنهم، بما في ذلك تصنيف الطلبة في مجموعات، أو ترتيبهم في شعب خاصة في المادة التعليمية، وتوجيه بعضهم نحو تعليم علاجي، وتعريضهم إلى دروس تقوية وغير ذلك.

ويقوم عدد غير قليل من المعلمين بالتشاور مع الطلبة في مناقشة خطته في التقييم من حيث عدد مرات التقييم ومواعيدها ووزن كل منها في الدرجة النهائية. عند ذلك فسيكون بمقدور الطالب تنظيم جهوده في ضوء هذه المعطيات. وفي فصول قادمة في هذا الكتاب، سيتم مناقشة عدة أمور تتعلق بتقدير الدرجات ومدى الاستفادة منها بالنسبة للمهتمين باختلاف أنواعهم. وكذلك تعميق قدرة المتعلم على الاستفادة من الأداء على التقييمات في تطوير مستواه.

إخراج الإختبار بشكله النهائي

يعد الشكل النهائي للتقييم مسألة هامة، وخاصة لأنها تتعلق بمسألة الصدق.إذ أن النسخة التي تحوي أخطاء مطبعية وأخطاء في ترتيب الصفحات ولغة صياغة الأسئلة، ستؤدي إلى إزعاج الطلبة الجادين وتزودهم بانطباع مفاده أن المعلم لم يقم بمسؤولياته المنوطة به جيدا.

توزيع فقرات الإختبار

كمبدأ عام، يفضل كتابة أو طباعة أي أداة تقييم (اختبار أو مقياس) يتطلب من الطلبة كتابة الإجابة عليه. كما يفضل أن يتم تزويد نسخة من هذا التقييم لكل مفحوص. وقد يلجأ بعض المعلمين إلى كتابة أسئلة التقييم على السبورة ، وهنا يجب مراعاة قدرة الطلبة على رؤية ما يكتب

بوضوح. كما قد يلجأ بعض المعلمين إلى قراءة الأسئلة على مسامع المفحوصين، وهنا يخشى أن لا يتمكن جميعهم متابعة ما يقرأ لسبب ضعف السمع أو عدم وضوح لفظ بعض الحروف أو الكلمات من قبل بعض المعلمين.

تصميم كراسة الإختبار

يتوقع أن تضم كراسة الإختبار الفقرات التي تتكون منها أداة القياس وكذا تعليمات الاجابة وإرشادات التصحيح. إضافة إلى بعض المعلومات الضرورية كعدد صفحات التقييم وأية معينات لتسهيل عملية الأداء على الإختبار كما في الجداول والأشكال والرسوم البيانية وغيرها. وتعرض البنود التالية وصفا لمكونات كراسة الإختبار.

تنظيم فقرات الإختبار:

إذا كان هناك عدد كبير من الفقرات الموضوعية، فيجب ترتيبها معا في كراسة خاصة، ويفضل أن يتم إخراج الأسئلة المقالية في ورقة مستقلة إذا كان الإختبار يتألف من نوعين من الفقرات. وبالنسبة للاختبار المقدم على مجموعة كبيرة نسبيا من المفحوصين، فيجب أن يكون هناك غلاف للاختبار ويتضمن عددا من التوجيهات والإرشادات للمفحوصين. كما ينصح بإخراج الأسئلة بحيث يتم وضع الفقرات من نفس النوع كل في مجموعة منفصلة عن الأخرى. أي يتم وضع فقرات الصواب والخطأ معا، وفقرات الاختيار من متعدد معا، وفقرات المزاوجة معا، وهكذا.

والجدير بالذكر، أن وجود صفحة الغلاف للاختبار يمنع المفحوصين الذين يستلمون كراسة الامتحان أولا من فتح الكراسة والبدء بالإجابة، لأنه من المفروض أن لا يجوز للمفحوص البداية في الإجابة إلا

بعد أن يؤذن له بذلك . كما أن مسألة ترتيب الفقرات كل مجموعات حسب النوع، يسهل على المفحوص الانتقال من فقرة إلى أخرى، إذ لا يستوجب ذلك تغيير نمط الإجابة عند الانتقال من فقرة من نوع الاختيار من متعدد مثلا إلى فقرة صواب وخطأ.

يعتقد البعض بأن ترتيب الفقرات من النوع الواحد من السهل إلى الأكثر صعوبة يقلل مستوى القلق عند المفحوصين. وهذا ربما يشجع الطلبة عموما على الإجابة وتوفير جهودهم للإجابة على الفقرات التي تحتاج جهدا أكبر. كما يعتقد البعض الآخر أن يتم ترتيب الفقرات حسب ما تمثله في المحتوى. وهذا لا يمنع من ترتيبها حسب مستوى الصعوبة (من السهل إلى الأصعب في المحتوى الواحد).

كما يفضل تشجيع المفحوصين ضرورة الرجوع إلى محاولة الإجابة على الفقرات التي تم حذفها لسبب أو لآخر.

إرشادات للطلبة: يفضل أن تتضمن إرشادات الأداء على التقييم (الإختبار) معلومات مختصرة. كما في عدد الفقرات، ونوعها، والوقت المحدد للاختبار، وكذلك توجيهات تتعلق بكيفية الإجابة وآلية وضع اختيار الجواب الصحيح. إضافة إلى المعلومات المتعلقة بإمكانية التصحيح لأثر التخمين، والمعلومات الأخرى المتعلقة بإمكانية اختيار الإجابة على فقرات معينة وبكيفية معينة. وهناك بعض من الفقرات التي يحتاج الإجابة عليها تعليمات خاصة مثل:

*. اقرأ الفقرة المبينة تاليا للإجابة على الفقرات 12، 13، 14، 15.

يتم عرض الفقرة هنا مثلا

*. الفقرات : 9، 10، 11 تعتمد على البيانات الواردة في الجدول التالي:

يتم عرض الجدول هنا

إخراج الإختبار

تصوير (سحب) الإختبار Duplicating: على الرغم من توافر آلات التصوير في معظم المدارس، إلا أنه من الضرورة التأكد دائما من جاهزية تلك الآلات، وخاصة قبيل موعد التقييم (الإختبار) بوقت كاف. كما يجب أن نحرص على أن يتم إخراج أوراق التقييم واضحة وكاملة وتتضمن الجداول والأشكال مفصلة وواضحة. ولا ننسى التذكير هنا بضرورة أن يتم مراجعة نسخة الإختبار لغويا وتنظيما قبيل تصويرها.

سرية الإختبار Test Security: وتعد هذه مسألة غاية في الأهمية، إذا تم تصوير الإختبارات من قبل جهة مركزية في الكلية أو المدرسة وبغياب صاحب الإختبار. عند ذلك لا بد من استشارة إدارة المدرسة بشأن توفير السرية اللازمة . أما إذا كان التصوير يتم من قبل صاحب المؤسسة التعليمية، فإنه يتحمل شخصيا مسؤولية ذلك.

تحضير التقييم (الإختبار) للتصحيح

قبيل البدء بتصحيح أوراق الإجابة، يجب التأكد من أن مفتاح التصحيح سليما . كما ينصح بالإسراع في إنجاز عملية التصحيح وإعلام المفحوصين بنتائجهم لتعزيز سلوكهم، والحصول على التغذية الراجعة المناسبة، وكذلك زيادة دافعيتهم. إن المراجعة المسبقة لمفتاح التصحيح تساعد الفاحص في استقصاء أي أخطاء في التقييم ليتم تفاديه.

بشكل عام، لا ينصح باستخدام ورقة إجابة منفصلة عن كراسة الأسئلة للطلبة في الصفوف الثلاثة الأولى. ولا ضير من استخدامها في حالة الصفوف المتقدمة، ليتسنى للفاحص استخدام كراسة الأسئلة لاستخدامات مستقبلية. وفي حالة التصحيح على نموذج الإجابة، فإنه من الممكن استخدام المفتاح المثقب أو آلة التصحيح الماسحة **Scanner** إن توفرت تلك الآلة والأوراق الخاصة بها. أما في حالة الأسئلة المقالية، فيتم تصحيحها بطرق أخرى.

أسئلة وتمارين

- ناقش العيوب التي تعاني منها الإختبارات من صنع المعلم. واذكر الخطوات التي تساعد على تلافي كل منها.
- ما العوامل الرئيسية التي يجب مراعاتها في مرحلة التخطيط لبناء الإختبار؟
- إذا أراد المعلم قياس قدرة طلبته على الاستدلال، فهل يمكنه استخدام الإختبار الموضوعي؟ لماذا؟
- هل تعتقد أننا سنستغني في المستقبل عن الإختبارات التحصيلية من صنع المعلم؟ دافع عن رأيك.
- هل يجب بناء الإختبارات التحصيلية في المدرسة من قبل معلم واحد، أم من قبل لجنة؟ ناقش إيجابيات وسلبيات كلا الطريقتين.
- كيف تبرر استخدامك للفقرات من النوع الانتقائي في الإختبارات لمجموعة من الآباء الذين يعارضون استخدامها.

أجب بـ "نعم" أو بـ "لا"

*. يتحدد الهدف الرئيس من إعداد جدول المواصفات في تحديد الزمن الأمثل للاختبار.

*. يهتم الإختبار محكي المرجع في الكشف عن التباين بين أداء المفحوصين.

*. يتركز الفرق بين الادوات محكية ومعيارية المرجع في شكل الفقرات المستخدمة.

*. تزداد صعوبة فقرة الاختيار من متعدد بزيادة تجانس المموهات.

*. أغلب الإختبارات المقننة معيارية المرجع.

*. يعد اختبار التمكن أحد أنواع الإختبارات معيارية المرجع.

*. يكون الإختبار جيدا إذا كانت درجات الطلبة عليه عالية.

*. تستخدم الإختبارات الصياغية في قياس أداء الأطفال في المرحلة الابتدائية.

*. يتم اشتقاق المموهات الجذابة للفقرات الانتقائية من احتمالات أخطاء الممتحنين.

*. المموه الجيد هو من تكون نسبة الاجابة عليه من الطلبة ذوي العلامات الكلية المنخفضة (ضعاف التحصيل) أقل منها عند أولئك مرتفعي الدرجات (المتفوقين).

*. يفضل استخدام الإختبارات المقننة بدلا من تلك التي من صنع المعلم بفضل ثباتها.

*. تتشابه أدوات القياس في التصميم والتطبيق.

*. ينصح بأن تتضمن الفقرة نصا حرفيا صحيحا منقولا من الكتاب المقرر.

ضع دائرة حول رمز الاجابة الصحيحة

*. الفائدة الرئيسة من استخدام جدول المواصفات في بناء أدوات القياس هي.......

أ) توفير الوقت

ب) تجعل فقرات الإختبار اكثر سهولة

ج) تزيد في موضوعية الإختبار

د) تحسن في تمثيل المجال السلوكي أوالمحتوى

*. ما أفضل أنواع الفقرات في حالة قياس قدرة الطالب على التصنيف؟

أ) الصواب والخطأ ب) الإجابة القصيرة

ج) المزاوجة د) الاختيار من متعدد

*. لتقليل تأثير عامل التورية، ما أفضل نوع من الفقرات التي ينصح باستخدامها؟

أ) المقالية ب) المزاوجة

ج) الاختيار من متعدد د) الإكمال

***. تأمل الفقرة التالية وبين جوانب الضعف فيها، ثم أعد صياغتها مراعيا لشروط الصياغة:**

الماء مركب كيماوي، يتألف من اتحاد عنصرين كيماويين هما الهيدروجين والأوكسجين بنسب متفاوتة نتيجة تفاعل كيماوي معقد، ويفيد الماء في شتى مناحي الحياة الصناعية والتجارية والطبية والغذائية، إذ لا يكاد يخلو نشاط بشري من استخدام الماء. والماء مالح وعذب. ما رمز مركب الماء؟

أ) الرمز هو **H2O** ب) الرمز هو **H2O2**

ج) الرمز هو **HNO3** د) الرمز هو **HOH**

***. تأمل الفقرة التالية وبين جوانب الضعف فيها، ثم أعد صياغتها فـــي الفراغ المقابل:**

جميع المدن التالية ليست ساحلية ما عدا ---.

أ) الرياض ب) العين ج) الفجيرة د) جبل العرب

***. تأمل فقرة المزاوجة التالية وبين جانب/جوانب الضعف في صياغتها:**

المقدمات	الاجابات
1- أمين عام الأمم المتحدة الحالي	أ. النيل
2- عاصمة بريطانيا	ب. أفرست
3- أطول نهر في العالم	ت. لندن
4- أعلى قمة جبل في العالم	ث. بانكي مون
5- أكبر الدول العربية سكانا	ج. مصر

*. تأمل الفقرة التالية وبين نقاط الضعف في صياغتها حسب الارشادات:

-------- يساوي مجموع --- ضلعي المثلث ------الزاوية.

الفصل الثالث

التصحيح وتقدير الدرجات

Scoring and Grading

- مقدمة
- التقييم
- التقييم الصفي
- الإفتراضات التي يعتمد عليها التقييم الصفي.
- الدرجات ما لها وما عليها
- طرق التعبير عن تقدم الطالب
  - قوائم الفحص
  - التقارير القصصية
  - مؤتمرات المعلمين وأولياء الأمور
- أنظمة الدرجات والعلامات المتعدد
- الإتجاهات نحو الدرجات والعلامات
- المهارات الأساسية الضرورية لأخذ التقييم
- حكمة الاختبار
- قلق الاختبار
- التصحيح لأثر التخمين
- قواعد التقدير

3

المخرجات التعليمية

بعد دراسة هذا الفصل يتوقع من الطالب أن يكون قادرا على أن

- يصف العلاقة بين التقييم المستمر وعملية تقدير الدرجات.
- يعدد الاستخدامات المختلفة للدرجات.
- يعلق على أنواع النقد الموجهة للدرجات والعلامات.
- يصف طرق التعبير عن تقدم الطالب في البرنامج الدراسي.
- يقارن بين طرق أنظمة العلامات المتعدد.
- يصف استراتيجيات حكمة الإختبار.
- يصنف الطلبة حسب درجة قلق الإختبار.
- يصمم قاعدة تقدير لمهارة يختارها.

مقدمة

أخبرني أحد زملائي (عندما كنت عميدا لكلية العلوم التربوية) أن طالبا قدم إليه معترضا على علامته التي حصل عليها في مادة دراسية معينة وكانت (B+)، وهي تمثل مجموع النقاط التي أحرزها ذلك الطالب في تلك المـــادة على مدى الفصل الدراسي.

وجدير بالذكر أن (90) هي العلامة التي تمثل نقطة الانتقال من (B+) إلى (A). من ناحية ثانية، كان ملف الطالب يشير إلى أنه كان قد أحـــرز (87) في امتحان منتصف الفصل، وحصل على (91) في الامتحان النهائي، كما حصل على (90) على الورقة البحثية التي أعدها كمتطلب لتلك المادة على مدى الفصل الدراسي، وكان معدل درجاتـــه علـــى الامتحانـــات اليوميـــة والقصيرة التي خضع لها يساوي (89).

بناء على جميع تلك المؤشرات، وباستخدام معادلة تقدير الدرجات Grading Formula التي أعدها زميلنا، حصل الطالب على معدل عام (89.25) فـــي هذه المادة الدراسية وبهذا فهو يستحق (B+).

يقول زميلنا، بأن الطالب كان يتوق إلى أن تكون درجتـــه علـــى الورقـــة البحثية التي أعدها (93) بدلا من (90). وبالتالي ستكون درجته في المادة الدراسية ككل (A). وبالتالي فهو سيرشح للحصول على بعثة دراسية خارج البلاد. إضافة إلى هذا، لم يكتف الطالب بالادعاء بأن ورقته البحثيـــة قـــد أعطيت درجة أقل مما كانت تستحق، بل ذكر بأنه:

- لم يتغيب مرة واحدة عن المحاضرات،
- وكان يشارك في أعمال الفصل،
- وغالبا يشارك في النقاش حيثما طلب منه ذلك.

يقول زميلنا (أستاذ المادة) بأنه كان مهتما بتشجيع طلبته على الحضور والمناقشة والمشاركة في الأعمال الصفية، ولكنه لم يضع مقابل ذلك نقاطا في معادلة تقدير الدرجات عند تقييم طلبة الصف. إذ لو أعطي هذا الطالب نقاطا على عدم تغيبه عن المحاضرات، أو لو كان يكافأ مقابل مشاركته في النقاش، فإنه سيحقق درجة (A) في المادة الدراسية وسيفوز بمسابقة البعثة الدراسية التي ربما تشكل نقلة نوعية في حياته المستقبلية.

وفي الختام، يبدو أن طريقة إعداد معادلة تقدير الدرجات Grading لتقييم أداء الطالب في الصف غاية في الأهمية، لذا فإن المفروض أن تكون موضوعية ما أمكن، لما يترتب على ذلك من قرارات وتبعات تتعلق بمصير الطلبة. الأمر، الذي يبدو مهما للغاية الاطلاع على طرق تقدير الدرجات (كمية أو نوعية) مع الأخذ بعين الاعتبار تصنيف الإختبارات حسب تفسير النتائج عليها (معيارية المرجع أو محكية المرجع).

من زاوية أخرى، حتى يكون تقييم أداء الفرد موضوعيا ودقيقا ومتوازنا، لا بد من أن يكون متنوعا ومتعدد المصادر، ومستمرا على فترة زمنية كافية.

التقييم: متعدد المصادر مستمر الإجراء ومتوازن

يعرف التقييم Assessment بأنه عملية يومية منظمة يمكن بواسطتها جمع البيانات عن مستوى التقدم في أداء الفرد (في التحصيل مثلا) (Nitko, 2005). وحتى يكون ذلك النشاط فعالا، لا بد له من أن يكون مستمرا ويتضمن خصائص كل من التقييم البنائي Formative Assessment والتقييم التشخيصي Diagnostic Assessment والتقييم الختامي Summative Assessment.

ومن الجدير القول بأن التقييم المستمر يستخدم في اتخاذ قرارات تتعلق بالتخطيط لدروس يومية، ومن ثم تطوير تلك الدروس. ومن الشائع

أيضا، أنه لا يتم تسجيل نتائج التقييم البنائي في سجل الطالب النهائي. بل يلجأ البعض إلى تزويد الطالب بتلك النتائج مباشرة. كما يمكن أن تناقش نتائج ذلك التقييم مع أولياء الأمور من خلال اجتماع خاص، أو عن طريق الاتصال التلفوني، أو بالمراسلة البريدية. وهذا يعتمد على الهيكل التنظيمي للمؤسسة التعليمية وإمكاناتها الإدارية والمادية.

تشير عملية تقدير الدرجات Grading مبدئيا إلى عملية استخدام نظام الرموز (عادة الحروف) للتعبير عن أشكال تقدم الطالب. ولأغراض ختامية، تساعد هذه العملية في تزويد المعلمين، وأولياء الأمور، والطلبة، والمهتمين بمعلومات عن مدى تحقيق الطالب للأهداف التعلمية. وهذه الدرجات التي تعطى للطالب في نهاية المرحلة الدراسية، يمكن أن تكون بمثابة كشف الدرجات الذي يستخدم في أغراض متعددة من قبل المؤسسات الأكاديمية العليا، أو هيئات الاستخدام والتوظيف.

من ناحية أخرى، فإن تلك الدرجات التي تستخدم في الإيفاء بعدد من الأغراض التقييمية لا بد لها من أن تستند على تقييم رسمي ومستمر ينسجم مع معايير المدرسة والأهداف التعلمية للمنهاج الدراسي. ولأن جمهور المهتمين أو مستخدمو الدرجات Stakeholders يستخدمون الدرجات النهائية لتحقيق عدد من الأهداف، فلا بد لها من أن تكون صادقة ومتعددة المصادر مثل:

- **المقاييس**
- **الإختبارات**
- **الامتحانات بأنواعها**
- **الأداء على المهام**
- **المشاريع البحثية**
- **ملفات الإنجاز**
- **المنتجات والأعمال التي يصممها أو يصنعها الطلبة**

إن من غير العدالة أن تعتمد الدرجة النهائية للطالب في مادة دراسية معينة على امتحان منفرد أو امتحانين، لأن هذا الأسلوب من التقييم يتجاهل بدرجة كبيرة عددا من العوامل الهامة في تقييم الطلبة مثل:

- **ربما يكون المتعلم على علم بعدد أكبر من المهام والمعارف من تلك التي يتضمنها تقييم واحد.**
- **تؤدي الظروف الشخصية والأسرية للطالب (كالمرض وغيره) إلى التدخل في قدرة الطالب على إظهار ما لديه من مستوى تحصيلي مطلوب.**
- **يستطيع الطالب أن يعبر عن مستوى تحصيله الدراسي في أكثر من أسلوب غير تلك التي يتطلبها التقييم الذي صممه المدرس لطلابه.**
- **هناك عدد غير قليل من الأهداف التعلمية التي يتطلب تقييمها من خلال مشاريع بحثية، وأوراق طويلة المدى. أو من خلال مهام وواجبات خارج إطار المدرسة.**

من ناحية أخرى، تعتبر عملية تقدير الدرجات إحدى أصعب جوانب العملية التدريسية بالنسبة للمعلمين. وهذه مقولة شائعة منذ زمن بعيد، ويعزي إيبل (Ebel, 1979) ذلك إلى الأسباب التالية:

- **ليس من السهولة تقييم التحصيل بالشكل المطلوب.**
- **إن وجود آراء تربوية وفلسفات مختلفة تستوجب استخدام طرق لتقييم وتقويم مستوى تقدم الطلبة.**
- **تتطلب عملية "تقدير الدرجات" من المعلمين الحكم على أداء الطلبة. ومعظم تلك العمليات صعبة وغير سارة التنفيذ.**

التقييم الصفي

في ضوء ما تقدم، تبرز أهمية الحديث عن التقييم الذي يتم تطويره واستخدامه وتصحيحه بواسطة معلم أو مجموعة من المعلمين بقصد تقييم أداء الطالب أو أداء الطلبة في وحدة دراسية، أو في مجال معرفي معين، وهذا يعرف **بالتقييم الصفي أو المدرسي** Classroom Assessment. ويفيد هذا النوع من التقييم في التأثير في عملية التدريس وزيادة فاعليتها، وكذلك مساعدة الطلبة في تحقيق مستوى التحصيل المطلوب (**Nitko, 2002; Trice, 2000**).

وبشكل أدق، يعرف التقييم الصفي بأنه مجموعة من آليات التدريس وتقنيات التعلم التي تعتمد على مقولة أنه كلما تستطيع أن تعرف أكثر عن كيفية تعلم الطلبة تكون أكثر قدرة على تخطيط الأنشطة الصفية لتوجيه عمليتي التعليم والتعلم.

من ناحية أخرى، يفترض أن تكون تقنيات التعلم تلك غاية في البساطة **simple**، وتحقق الشمولية **anonymous**، وليست خاضعة للدرجات **non-graded** ، وتكون بمثابة نشاطات صفية **in-class activities** تزود المعلم والطالب بمراجعة مفيدة حول طرق التعليم والتعلم.

الإفتراضات التي يستند عليها التقييم الصفي

يختلف التقييم الصفي عن الإختبارات tests وبعض جوانب تقييم الطالب الأخرى، في أنها تهدف إلى تحسين عمليتي التعليم والتعلم أكثر من تحديد الدرجات. وتتركز الغاية العظمى من التقييم الصفي في الفهم الأفضل لتعلم الطلبة، كي يؤدي ذلك إلى تحسين في طرق وفلسفة التدريس نفسها أو الأهداف أو المحتوى. ويعتمد التقييم الصفي على عدد من الإفتراضات التالية:

- تعتمد نوعية تعلم الطلبة بدرجة كبيرة على نوعية التعليم. لذلك فإن أفضل طريقة واعدة في تحسين التعلم هي أن نبدأ بتحسين التعليم.
- لتحسين فعالية أداء المعلمين **Effectiveness** يتطلب منهم تحديد غاياتهم وأهدافهم بوضوح، ليسهل الحصول على تغذية راجعة وشاملة تساعد في تحقيق تلك الأهداف.
- لتطوير تعلم التلاميذ، يستوجب تزويدهم بتغذية راجعة مركزة ومتكررة وبالوقت المناسب. كما أنهم بحاجة إلى أن يتعلموا كيف يتم تقييم تعلمهم الذاتي. أي أن دور التقييم يهدف إلى إعطاء الدرجات النهائية لمساعدة تقدم الطلبة نحو تحقيق الهدف المنشود أو كلاهما.
- إن التقييم الذي من المتوقع أن يحسن التعليم والتعلم هو الذي يتم تصميمه من قبل المعلمين بقصد الإجابة عن الأسئلة التي يتم طرحها استجابة للمسائل والقضايا التي تبرز أثناء تدريسهم. كما يتوقع أن تعمل نتائج التقييم المؤسسي باتجاه تصميم وتطوير المنهاج أو تنظيم البرامج.
- إن البحث المنظم والتحدي الذكي هما مصدران رئيسان للدافعية، والنمو. والتقييم الصفي هو من يبرز هذا التحدي ويعظمه.
- يحتاج التقييم الصفي تدريبا مركزا، لأنه يستلزم معلمين محترفين ومنتمين من كافة الحقول والتخصصات.
- بمشاركة الزملاء المعلمين وإشراك الطلبة في فعاليات التقييم الصفي، يعمل هؤلاء على تطوير التعلم ورفع درجات الرضا الشخصي. إذ أن التقييم الصفي نشاط اجتماعي وجماعي.

تتطلب عمليتا التعليم والتعلم من المدرس أن يجمع مجموعة من البيانات بشكل مستمر ويتخذ في ضوئها عددا من القرارات بواقع قرار كل دقيقتين أو ثلاثة دقائق **(Shavelson & Stern, 1981)**. فالعملية التدريسية عموما

تسير وفق خطوات منظمة تبدأ بتحديد الأهداف التي تشير إلى نواتج التعلم أو المخرجات التي يتوقع من المتعلم اكتسابها نهاية العملية. فهي توجه عمل المعلم وتساعد في نجاح عمليات التقييم، وفي اتخاذ قرارات للدخول في مراحل جديدة.

وقد قدر الباحثون أن المعلمين يبذلون ثلث أو نصف وقتهم في نشاطات ذات علاقة بالتقييم (**Stiggins, & Associates, 1992**). وفيما يلي عدد من الأسئلة التي من المفروض أن يتم الإجابة عليها من قبل المعلمين عند اتخاذ قرارات **قبل وأثناء وبعد** عملية التدريس.

أ. **قرارات قبيل البدء بالتدريس:** إذ يجب على المدرس أن يطرح الأسئلة التالية:

i. ما المحتوى الذي يجب تغطيته خلال اليوم، الأسبوع، الشهر التالي؟ وهذا يتطلب من المعلم مراجعة محتوى المنهاج، ومخطط المساق، وعناصر الكتاب المقرر، والامتحانات التي يجب على الطلبة النجاح فيها.

ii. ما القدرات التي يتمتع بها الطلبة، والتي يجب أخذها بعين الاعتبار من قبل المعلم لتخطيط الأنشطة التدريسية؟ وهذا يتطلب من المعلم ملاحظة الطلبة غير الرسمية في الصف، التحدث مع معلمي الطلبة السابقين، الاطلاع على نتائج الطلبة على اختبار استعداد استخدم من قبل معلمي الطلبة السابقين، التعرف على المستوى السابق لتحصيل الطلبة ودرجاتهم على الإختبارات المدرسية الرسمية والمعلومات الشخصية الأسرية للطالب.

iii. ما المواد التعليمية التي يمكن أن يستخدمها المعلم وتناسب الطلبة؟ وهذا يتطلب ملاحظة المعلم المستمرة اليومية للتعرف

على مستوى دافعية الطلبة، واهتماماتهم، ومعتقداتهم، واتجاهاتهم نحو تعلم بعض الموضوعات، الاطلاع على نتائج الإختبارات القبلية، وتقييم المعلمين السابقين لهم، والنتائج على الإختبارات التحصيلية المقننة. كما يتطلب خبرة كافية من المعلم بالمحتوى الذي سيدرس.

iv. ما الأنشطة التعليمية التي من المفروض أن ينخرط بها المعلم وطلبته عند تدريس الوحدة أو المساق المقرر؟ وهذا يتطلب معرفة أنواع الأنشطة التي استخدمها المعلم وعملت على تحسين اهتمامات الطلبة، وتحليل تتابع الأنشطة التعليمية التي يكلف الطلبة القيام بها.

v. ما الأهداف التي يريد المعلم أن يحققها طلبته نتيجة تعليمه لهم؟ وهذا يتطلب مراجعة الغايات targets والأهداف التعلمية **learning objectives** ، ومراجعة الأسئلة التي سيكون الطلبة قادرين على إجابتها، ومراجعة أنواع الأداء ومهارات التفكير المتوقع منهم إظهارها بعد التعلم.

vi. كيف يمكن للمعلم أن ينظم ويرتب الطلبة في غرفة الصف للدروس القادمة ؟ وهذا يتطلب مراجعة سجلات الطلبة ودراسة أحوالهم وخصائصهم من خلال الملاحظة اليومية والمستمرة لسلوك الطلبة في الصف وخلال الفترات السابقة، والتعرف على أفضل تصنيف للطلبة في الصفوف. وبما يسهل عملية تعلمهم أثناء العملية التدريسية.

ب. قرارات أثناء عملية التدريس:

i. هل يسير الدرس بالاتجاه الصحيح ؟ وهل أن الطلبة متابعون بدرجة جيدة؟ وهذا يتطلب من المعلم ملاحظة الطلبة خلال

الأنشطة التعلمية، وتحليل استجاباتهم نحو الأسئلة التي يتم طرحها، إضافة إلى دراسة مناقشات الطلبة في الفصل.

ii. ما الذي يجب أن يعمله المعلم لضمان سير الدرس نحو الأفضل؟ وهذا يتطلب تشخيص أخطاء الطلبة، واستمرار البحث عن طرق بديلة لعرض المادة التعليمية. وكذلك العمل على تحديد الطلبة ذي المشاركة المحدودة.

iii. ما التغذية الراجعة التي من المفروض أن يقدمها المعلم لكل طالب؟ وهذا يتطلب من المعلم استخدام ملاحظاته المستمرة وخبراته السابقة لتحديد نوع ومقدار التعزيز والتقدير للطلبة، وكذلك استخدام المعلومات المتوفرة عن أداء الطالب على الامتحانات القصيرة وفي المقابلات.

iv. ما مدى مقدرة الطلبة وجاهزيتهم للانتقال إلى النشاط التعلمي اللاحق ؟ وهذا يتطلب من المعلم القيام بملاحظة أداء الطلبة في الامتحانات النهائية والواجبات المنزلية، والتأكد من مدى فهم الطلبة لتلك الواجبات.

ت. قرارات بعد عملية التدريس:

i. إلى أي مدى حقق الطلبة الأهداف التدريسية بعيدة المدى وقصيرة المدى؟ ويمكن التحقق من ذلك من خلال النتائج على كل من الامتحانات الصفية، والمشاريع، والملاحظات، والمقابلات، وكذلك النتائج على الإختبارات المقننة.

ii. ما نقاط القوة ونواحي الضعف التي يجب إبرازها وتوضيحها للطلبة وأولياء أمورهم؟ وهذا يتطلب معرفة المعلم بمشاركة الطالب في الأنشطة الصفية، ومراجعة الواجبات المنزلية،

وتحليل الأداء على الإختبارات، ومراجعة البيانات المتعلقة بالخلفية الأسرية للطالب.

iii. ما الدرجة التي يستحقها كل طالب على الوحدة الدراسية؟ وهذا يعتمد على مستوى الأداء على الامتحانات القصيرة، ومدى التقدم الذي حققه الطالب على المشاريع البحثية الصفية، والأنشطة التعلمية، وكذلك الملاحظات اليومية في المدرسة وخارجها.

iv. إلى أي حد كان المعلم فاعلا **effective** في تدريس هذه المادة للطلبة؟ وهذا يتم من خلال مراجعة متوسط أداء طلبة الصف على الأهداف التدريسية المهمة، ومستوى أداء الطلبة على بعض الأسئلة التي تضمنها بعض الإختبارات المقننة، وكذلك من خلال تقييم مدى تعلق الطلبة وحبهم بالأنشطة الصفية والمواد التعلمية.

v. ما مدى فعالية المنهاج والمواد التي تم استخدامها؟ ويتم تقييم ذلك من خلال مراجعة الملاحظات المتعلقة بمدى اهتمام وتفاعل الطلبة تجاه المواد والأنشطة التعلمية، مطالعة مستوى تحصيل الطلبة على الإختبارات الصفية التي تقيس المنهاج، وكذلك مراجعة أداء الطلبة على بعض الوحدات المختارة من الإختبارات المنتظمة والمخطط لها والتي تجري في المدرسة.

طرق التعبير عن تقدم الطالب

تنوعت طرق ووسائل التعبير عن تقدم الطالب في برنامجه الدراسي نظرا لتنوع وتداخل مجالات النمو الإنساني واختلاف درجات تعقيدها وشمولها من ناحية، وبناء على اختلاف مزايا وفعالية أو محددات أساليب

التعبير عن مدى ذلك التقدم من ناحية أخرى. وبشكل محدد، فقد يتم استخدام بعض اختبارات كل من الذكاء بأنواعه والتحصيل، وكذا المقاييس النفسية وقوائم الشخصية وأساليب الملاحظة أو المقابلة في تقييم مدى التقدم. وعليه، تختلف طريقة التعبير عن هذا التقدم باختلاف أسلوب التقييم المتبع. إذ يمكن التعبير عن واقع الطالب **كميا أو نوعيا.**

تقدير الدرجات نوعيا Qualitative Grading ويكون هذا عندما يتم التعبير هجائيا أو لفظيا عند تقييم مستوى أداء الطالب ومدى تقدمه. كأن يتم التعبير عن ذلك بالحروف التالية:

- E: ويشير إلى أن مستوى الأداء ممتاز Excellent،
- S: ويشير إلى أن مستوى الأداء مرضي Satisfactory،
- NS: ويشير إلى أن مستوى الأداء غير مرضي Not Satisfactory،
- NI: ويشير ذلك على أن الأداء يحتاج تطويرا Needs Improvement

تقدير الدرجات كميا Quantitative Grading عندما يتم التعبير عن مستوى الطالب رقميا كما في الحروف (F, E, D, C, B, A,) حيث تشير هذه الحروف إلى دلالات رقمية.

وسيتم التعرض في هذا الفصل إلى عدد من طرق التعبير عن مستوى تقدم الطالب المتمثلة بالإختبارات، والمقاييس، وقوائم الفحص (الشطب)، والتقارير القصصية بأنواعها، ومؤتمرات المعلمين وأولياء الأمور.

قوائم الفحص Checklists: هي مجموعة من العبارات التي تعبر عن سلوكات معينة يقوم المعلم أو المقوم بفحص مدى توفرها في المفحوص خلال فترة زمنية معينة (عام دراسي أو فصل دراسي أو غيرها). وفي الجدول التالي، عرض لقائمة فحص لتقييم قدرة طالب في

المرحلة الثانوية على التحدث speaking تم ترجمته بتصرف من نيتكو (Nitko, 2005).

قائمة فحص حول قياس قدرة طالب ثانوي على المحادثة

سلوك النطق (المحادثة) Speaking Behavior		التاريخ			إشارة موافقة مع التعليق
		اليوم	الشهر	السنة	
يتكلم بصوت واضح ومسموع.					
يتكلم لغته الأم بطلاقة.					
يعبر عن الأفكار بجمل كاملة ومفيدة.					
يستخدم عبارات وجمل مناسبة.					
يختار مواضيع مناسبة لعرضها.					
ينظم المادة	أ. يعرض عبارات الافتتاح والخاتمة.				
	ب. يستخدم ملاحظات أو بطاقات ملاحظة.				
	ج. يستخدم وسائل بصرية ومواد مساعدة.				
	د. يستخدم وسائل أخرى......				
يستخدم التواصل بالنظر ليضمن انتباه المستمع.					
ينوع من لهجته ونبرة الصوت وشدته لإيصال المعنى.					
يقف باعتدال منتصباً عند الكلام.					
يظهر ثقة بالنفس.					
يستخدم تعبيرات و لغة مناسبة لإيصال المعنى.					

يستخدم لغة مناسبة الموقف و التواصل.				
يركز على الأفكار الرئيسية.				
يستخدم أساليب الإقناع.				
يوصل الغرض المراد إيصاله عند الكلام.				

التقارير القصصية (السردية) Narrative Reports

يستخدم بعض المعلمين أسلوب إنشاء وكتابة التقارير عند تقييم أداء الطلبة في موادهم الدراسية، وخاصة خلال المرحلة الثانوية وأحيانا في المرحلة المتوسطة. والقليل منهم يستخدم التقارير القصصية في المرحلة الأساسية. وتختلف اتجاهات المعلمين نحو استخدام طرق التقدير، فهم لا يحبذون التقدير بالتقارير لكونها تستغرق وقتا طويلا لإعدادها، وبعضهم غير مقتنع بالتقييم باستخدام الدرجات والعلامات وطرق حسابها. والحقيقة أن اتباع هذا النظام أو ذاك هو قرار المدرسة وليس أمرا يقرره المعلم.

من ناحية أخرى، تنبع أهمية التقييم من خلال التقرير الذي يصف مجمل أعمال الطالب من خلال ما يقدمه من معلومات للطلبة أنفسهم، ولأولياء الأمور، وللمدرس الذي سينتقل الطالب إليه في المرحلة القادمة، مقارنة مع التقييم بالحرف أو الرقم. وبشكل أدق يتوقع أن يتضمن التقرير النقاط التالية:

- وصف تفصيلي لتحصيل الطالب مبينا نواحي قوته ومواطن ضعفه.
- وصف لسلوك الطالب أثناء التعلم، وكيفية تفاعله مع المادة الدراسية، وطريقة تعامله ومدى حبه للمادة، وكذلك كيفية تعامله مع أقرانه والعمل مع الفريق.
- مجمل توجيهات تضمن استمرارية وتحسن تعلم الطالب.

وهذا مثال لتقرير عن مستوى أداء طالبة في الصف الرابع في مادة التهجئة Spelling مقتبس بتصرف من ترايس (Trice, 2000):

أظهرت "سعاد" تحسنا ملحوظا في قدرتها على التهجئة وهي بمستوى أداء صفها بناءا على النتيجة التي حصلت عليها في امتحان طبق لهذا الغرض قبيل ثلاث أسابيع.

كانت "سعاد" تبذل جهدا ملحوظا في التحضير للامتحان، بالرغم من أنها مازالت تستخدم تهجئة خاصة Invented spelling عند استخدامها الحاسوب. كذلك، فإنها ما زالت في مستوى أولي في عمليات نقل ما تتعلمه من دروس التهجئة إلى الواجبات الكتابية.

من ناحية أخرى، بدأت تلك الطالبة وبنجاح استخدام الحاسوب في برامج ضبط التهجئة Spell Checker . وبشكل محدد، يبدو أنها بحاجة إلى تدريب على قاعدة (i) قبل (e).

يتوقع أن يفرح والدا "سعاد" كثيرا بهذا التقرير، كما أن سعاد ستعرف ما المطلوب منها عمله لتصويب أوضاعها. وإذا استخدم التقرير نهاية العام الدراسي، فسيكون مفيدا جدا للمدرس في السنة القادمة ليتمكن من التخطيط لتدريس سعاد على نحو أفضل. وبالرغم من الفائدة التي توفرها التقارير، إلا أنها تعاني من عدد من نقاط الضعف من مثل:

1. يستغرق كتابة التقرير وقتا طويلا نسبيا. إذ أن المطلوب من كل معلم أن يكتب عددا من هذه التقارير لكل طالب وفي كل مادة

دراسية وأربع أو ست مرات خلال السنة الدراسية. فالمعلم الذي يدرس صفا دراسيا يتألف من 25 طالبا ويدرسهم ست مواد مثل التهجئة والقراءة والكتابة والعلوم الطبيعية والرياضيات والعلوم الاجتماعية يقوم بكتابة عدد من التقارير بما يساوي كتابة رواية تقع في (300) صفحة.

2. يخشى من أن تكون معظم التقارير التي يكتبها المعلمون تركز على جهد الطالب أكثر مما تركز على تحصيله الدراسي بدقة.
3. يخشى أن يستخدم المعلمون عبارات معينة تتكرر في تقارير كافة الطلبة مع إظهار بعض الاختلافات الطفيفة في تقرير كل منهم.
4. يخشى أن يقترف بعض المعلمين بعضا من الأخطاء الإملائية أو التعبيرية عند توجيهها لأولياء أمور الطلبة. الأمر الذي يضعف مصداقية المعلم وهيبته أمام أسرة الطالب.

بناءا على ذلك، يبدو أنه من المهم تطوير المعلم بإرساله في دورات خارج منطقته كي يكتسب الخبرة المعمقة والقدرة على رؤية الأشياء وفقاً لعلاقاتها الصحيحة وأهميتها النسبية. كما أنه قد تكون هناك حاجة إلى التمرين نحو كتابة ملاحظات مفيدة وعبارات غير تهديدية، لا تنطوي على اللوم للطالب.
ولأن التقارير القصصية مطولة، وتستهلك وقتا أثناء إعدادها، قامت بعض المدارس بتعديل تلك التقارير، وعرفت بالتقارير القصصية المعدلة **Modified Narrative Reports**، وذلك بدمج كشف الدرجات مع كتابة بعض التعليقات والملاحظات المختصرة عن كل طالب.

مؤتمرات المعلمين وأولياء الأمور

تعد اجتماعات المعلمين وأولياء الأمور **Parent-Teacher Conferences** من أفضل الطرق لبناء علاقات قوية مع أولياء أمور الطلبة، بقصد تزويدهم

بنقاط قوة ومواطن ضعف أبنائهم، ولمساعدتهم الإنخراط بتعلم أبنائهم. من ناحية أخرى، لا بد من إدارة هذه الاجتماعات بحذر ومهارة عالية لضمان درجات نجاح عالية.

ومن محددات مؤتمرات المعلمين أنها تحتاج وقتا أطول للمعلمين في الإعداد لها وفي التنفيذ. فقد يستغرق المؤتمر الواحد عددا من الساعات وعلى حساب اليوم الدراسي. أو ربما يتم الاجتماع بعد وقت الدوام المدرسي خلال المساء بما يتفق مع ظروف أولياء الأمور. وعندما لا تتوفر الثقة في أطراف المؤتمر، يتوقع أن ترتفع فرص الإحباط والقلق للآباء والمعلمين. وقد يتعذر حضور العدد الكافي من أولياء الأمور للاجتماع بسبب المرض، أو الانشغال، أو الفقر، أو التعب، أو ضعف اللغة أو المستوى التعليمي لديهم. ولضمان نجاح تلك المؤتمرات وزيادة فاعليتها، لا بد من الانتباه جيدا للمراحل التالية (أ، ب، ت، ث، ج) وتفعيل محتوى كل منها ما أمكن:

أ. **التخطيط للمؤتمر** ويتم ذلك من خلال النقاط التالية:

1. ضرورة تحديد الأهداف التي ترغب إدارة المدرسة في تحقيقها.
2. إعداد جدول أعمال للمؤتمر بالتعاون مع أولياء الأمور.
3. إعداد رسالة موجهة لأولياء الأمور توضح أن اجتماعا سيعقد وتبين مكونات جدول الأعمال، وتحدد توقيتا محتملا لذلك الاجتماع.
4. تحضير الأسئلة المنوي طرحها خلال الاجتماع ، والنقاط المنوي التركيز عليها، والاقتراحات المنشودة، والاستراتيجيات المتوقع استخدامها خلال الاجتماع.

5. الطلب من أولياء الأمور إعداد قائمة بحاجات الطلبة ونقاط قوتهم.
6. إعداد ملف إنجاز لطالب ما كمثال، لتدعيم وجهة نظرك.
7. إدارة الوقت بالطريقة التي تسمح بمناقشة القضايا المطلوب بحثها والإجابة عن الأسئلة الرئيسة.

ب. تنظيم آلية الجلوس في الاجتماع Create the Right Setting

وذلك بـ:

1. استخدام مقاعد مريحة تلائم كبار السن.
2. الترحيب بالضيوف في مدخل المدرسة والجلوس معهم .
3. إعداد بعض المنشورات والوثائق (مثل: الهيكل التنظيمي لإدارة المدرسة، نظام التقييم، نشاطات مجلس المعلمين وأولياء الأمور) على طاولة الاجتماع، ليتمكن أولياء الأمور من حملها بعد الاجتماع.

ت. عقد المؤتمر Running the Conference ويشمل ذلك إتباع التالي:

1. الابتداء بمراجعة جدول أعمال الاجتماع مع الحضور.
2. ضرورة الإصغاء لأسئلة أولياء الأمور، ومن ثم الإجابة على أسئلتهم مباشرة وبوضوح. مع أهمية الطلب من أولياء الأمور بيان انطباعاتهم نحو نقاط الاهتمام المطروحة.
3. العمل بالتشارك مع الوالدين لتحديد أهداف محددة للطالب.
4. الحرص على أن يختتم المؤتمر باقتراح خطوات محددة ينبغي على أولياء الأمور القيام بها.

5. تكرار الدعوة لأولياء الأمور كي يزيد تواصلهم مع المدرسة بالوسائل المتاحة كالهاتف، والمراسله، أو حتى تبادل الزيارات.

ث. بعد المؤتمر After the Conference ويتم القيام بما يلي:

1. كتابة وتوثيق ما قيل في المؤتمر وحفظه في ملف خاص.
2. تصنيف القرارات والمداخلات التي انبثقت عن المؤتمر.
3. المتابعة خلال الهاتف أو المراسلة.
4. توثيق ما تم تعلمه في المؤتمر ليساعد المعلمين في تدريسهم.
5. القيام بتقويم المؤتمر وتحديد طرائق تحسين الاجتماعات القادمة.

ج. اعتبارات وتحذيرات Cautions and Considerations

1. ضرورة التعاطف مع غضب أولياء الأمور أو تشككهم، إذ ينبثق هذا من اهتمام الآباء بالطلبة.
2. إشراك الطلبة في الإعداد لتلك المؤتمرات (خاصة عند اختيار موضوعات معينة أو اختيار الأمثلة لعرضها في الاجتماعات).
3. لا ينصح باستخدام المؤتمرات فقط في إبلاغ أولياء الأمور عن مشكلات الطلبة، بل المفروض أن يتم التواصل مع المنازل بوسائل أخرى مثل المراسلة، والاتصال الهاتفي، والواجبات المنزلية. وكذلك ابلاغهم عن انجازات الطلبة وبيان مدى تقدمهم في كافة مجالات العملية التعليمية في المدرسة.
4. المرونة في تحديد مواعيد الاجتماعات مع أولياء الأمور الموظفين.

5. الانتباه إلى التغيرات التي تطرأ على التركيب الأسري لأهالي الطلبة، لمراعاة ما يطرأ على أسر الطلبة من حالات طلاق أو انفصال وغيرها.
6. مراجعة التفسيرات المناسبة للدرجات على الإختبارات خلال التحضير للمؤتمر وبالتنسيق مع مدير المدرسة والأخصائي النفسي.
7. التوضيح لأولياء الأمور أن الإختبارات المقننة ليست المصدر الوحيد للمعلومات.
8. ضرورة الدقة والحذر عند إبلاغ أية معلومة تتعلق بالطالب. إذ يجب أن تنسب المعلومة لمصدرها (معلم المادة، أو الأخصائي، أو الممرض).
10. ضرورة شرح عملية تقدير الدرجات بنوع من التفصيل.

قواعد التقدير Scoring Rubrics

قواعد التقدير هي قواعد تصحيح أو تقدير لأعمال الطلبة الأدائية، كما تستخدم في تصحيح استجابات الطلبة على فقرات الإختبارات الصياغية (الإنشائية). وهذه القواعد، تحوي مجموعة من معايير الأداء لمهام الأداء بعامة، وخاصة تلك التي:

1. تكون معقدة وذات أبعاد متعددة.
2. يمكن تقسيمها إلى أجزاء معينة.
3. تمتلك خصائص أو صفات يمكن وصفها على مدى معين.

تكمن أهمية استخدام قواعد التقدير في كونها ترتبط بتعليم الأفراد وأنماط تعلمهم. إنها على درجة كبيرة من الأهمية في توضيح مستوى أداء الأطفال

لهم ولأولياء أمورهم. من ناحية أخرى، تستخدم قواعد التقدير في تقييم مجموعات الأنشطة، والمشروعات، والتقديم الشفهي، وفي فعاليات التقييم الصفي. كما أن مكان وتوقيت استخدام قواعد التقدير لا يعتمد على مستوى الطلبة (الصف الدراسي) أو الموضوع الدراسي، ولكن على الهدف من التقييم.

تساعد قواعد التقدير الطلبة في تزويدهم بالتغذية الراجعة من أجل تطوير أدائهم. وكذلك، فهي تدعم الامتحانات إلى الحد الذي تشير إلى أن محكات محددة قد تحققت. وفي حال تحقق هذين الهدفين، فإن ذلك يشير إلى أن قواعد التقدير هي أساليب مناسبة للتقييم.

أنواع قواعد التقدير Types of Rubrics

هناك نوعان رئيسان من قواعد التقدير هما قواعد التقدير الكلي وقواعد التقدير التحليلي. وهناك نوع ثالث يمزج النوعين معا. وعند استخدام نوع دون آخر، فهذا لا يعني أن هذا النوع يلغي الحاجة أو الأهمية لاستخدام النوع الآخر. وسيتم تاليا تناول كل منها بنوع من التفصيل:

قواعد التقدير الكلي Holistic Rubrics: هي قواعد التقدير التي يتم من خلالها تقدير الأداء على المهمة (الناتج أو الإجراء) بشكل عام، دون تقييم أجزاء تلك المهمة بشكل منفصل. ويتمتع هذا النوع **بمزيتين** هما (1) يتم تقييم الأداء في ورقة المتعلم (المفحوص) بشكل أسرع من الطريقة التحليلية، (2) تتيح الفرصة أمام المعلم أن يعرض الورقة كعمل متكامل. من زاوية أخرى، يؤخذ على تلك الطريقة أمران هما: (1) تعطي درجة كلية شاملة ولا تعطي تفصيلات ربما تفيد الطالب أكثر لتحسين عمله، (2) تبرز تحيزا ما تجاه مسألة معينة (مثل المظهر الخارجي) للورقة. وبالتالي، فقد يؤثر ذلك على موضوعية Objectivity الدرجة المعطاة.

قواعد التقدير التحليلي Analytic Rubrics: هي قواعد تقدير للأداء على أجزاء المهمة (الناتج أو الإجراء) وتنتهي بتقدير درجه نهائية. وتبدو تلك الطريقة أكثر صدقا لكونها توفر فرصة تقييم أبعاد الأداء ومكوناته وقدرة المتعلم على إحداث تكامل بين تلك الأبعاد. **هذه القواعد هي المستخدمة في مجال التقييم التشخيصي.**

وبنفس السياق، هناك ثلاث ميزات لهذا النوع من القواعد هي: (1) توفر إمكانية تزويد الطالب بتغذية راجعة تعرفه نقاط قوته ونقاط ضعفه على كل مكون، (2) توفر فرصة التعمق والاطلاع على أداء الطلبة على كل مكون، ليتسنى معرفة الأجزاء التي أخفق فيها نسبة كبيرة من الطلبة ليصار إلى إعادة عرضها وتدريسها بعمق أكبر، (3) تساعد في التعرف أكثر على إعادة بناء قاعدة توزيع الدرجات على المكونات بموضوعية أكبر.

من ناحية أخرى، يؤخذ على تلك الطريقة ثلاث مآخذ هي: (1) يتم تقدير الأداء في ورقة الطالب ببطيء مقارنة مع ما يتم باستخدامه في حالة الطريقة الشاملة، (2) في بعض الأعمال الإنشائية، يصعب وضع درجات معينة على أجزاء يصعب تحديدها، (3) ربما يضيق المعلمون حديثو التعيين ذرعا بمقدار الوقت اللازم لبناء قاعدة تقدير تحليلي. والشكل التالي يعرض مقارنة بين خطوات بناء نوعي تلك القواعد:

تصميم قواعد التقدير Rubrics

خطوة (1): تفحص الأهداف التعلمية التي من المتوقع تحقيقها من خلال مهمة الأداء.

خطوة (2): تحديد وتعريف الخصائص attributes القابلة للملاحظة التي تريد أن تراها أو لا تراها في أداء المتعلم.

خطوة (3): كتابة وصف كل من هذه الخصائص.

قاعدة كلية Holistic Rubrics

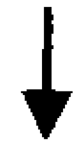
خطوة (4) أ: كتابة وصف تفصيلي للعمل المتميز والعمل الرديء الذي يتوقع من المتعلم إظهاره ويتم التعامل هنا مع الخصائص الفرعية بشكل كلي.



خطوة (5) أ: يتم وصف المستويات الأخرى على متصل التقدير الذي يمتد بين أداء متميز إلى أداء رديء.وذلك للأداء ككل.

قاعدة تحليلية Analytic Rubrics

خطوة (4) ب: كتابة وصف تفصيلي للعمل المتميز والعمل الرديء الذي يتوقع من المتعلم إظهاره على كل خاصية من الخصائص المبينة سابقا.



خطوة (5) ب: يتم وصف المستويات الأخرى على متصل التقدير الذي يمتد بين أداء متميز إلى أداء رديء وذلك لكل خاصية.

خطوة (6): إجمع عينات من أعمال الطلبة لتمثيل كل مستوى.

خطوة (7): أعد كتابة وتطوير قاعدة التقدير.

قواعد التقدير التحليلي الشامل Annotated holistic Rubrics: في هــذه الحالة، يتم استخدام التقييم الشامل أولا، ثم يتم تقدير الأداء على كل مكون من مكونات المهمة المنوي تقييمها. من زاوية أخرى، يصف الجدول التالي نوذج لقاعدة تقدير تقيم مكونات القاعدة عموما:

نموذج قاعدة تقدير تصف " مكونات قاعدة التقدير"

مستوى المهمة	يحقق الهدف تماما	يحقق الهدف تقريبا	لا يحقق الهدف
اختيار المهمة	المهمة مركبــة وتتطلــب مهارات تفكير عالية.	المهمة مركبة إلى حد مــا وتتطلب مهــارات تفكيــر متوسطة المستوى	المهمة بــسيطة وتتطلــب مهارات تفكيــر بمــستوى منخفض.
صياغة الأهداف	الأهداف واضحة وواقعية.	الأهداف واضحة إلى حــد ما وتبدو غامضة إلى حد ما.	الأهداف غير واضحة وغير واقعية.
مدى ونوعية الخصائص	وصـــفت الخـــصائص والصفات جيدا على مدى القدرات.	وصـــفت الخـــصائص والـــصفات علـــى مـــدى القدرات.	لم يتم وصف الخـصائص والـــصفات علـــى مـــدى القدرات.
قابلية الخصائص للملاحظة والقياس	قابلــة تمامــا للملاحظــة والقياس.	قابلة تقريبــا للملاحظــة والقياس.	غيــر قابلــة للملاحظــة أو للقياس.
مناسبة المهمة للفئة العمرية للمستهدفين	الخـــصائص والـــصفات ملائمـــة تمامـــا للفئـــة العمرية وقابلة للتطور.	الخـــصائص والـــصفات ملائمة نوعا مــا للفئــة العمرية.	الخصائص والصفات صــعبة جدا أو سهلة جــدا بالنــسبة للمستهدفين (غير مناسبة).
استخدام اللغة	صياغة الكلمات واضــحة ومفهومة تماما للمعنيــين (معلمين وآباء وطلبة).	صياغة الكلمات واضــحة ومفهومة بشكل تقريبــي للمعنيين.	صياغة الكلمات غير واضحة وغير مفهومة للمعنيين.
استخدام عبارات إيجابية	صـــيغت الخـــصائص كالأهداف بشكل إيجابي.	صيغت الخصائص إلى حد ما بشكل إيجابي.	صيغت الخصائص بــشكل لا يساعد في تنفيذها.

تمرين: إنتاج قاعدة تقدير Creating Scoring Rubrics

إذا اتجهت النية لتصميم قاعدة تقدير لتقييم أداء طلبة صف دراسي معين في مجال "مهارة الكتابة" أو "مهارة حل المسألة"، فإن المفروض أن يتم اتباع الخطوات التالية:

- **اشتقاق وتطوير مؤشرات الأداء:** إذا كنت خبيرا في ملامح ومواصفات الأداء (الكتابة، أو حل المسألة الرياضية مثلا)، بمستواه الأمثل، فما عليك إلا أن تكتب هذه الصفات والملامح (المحكات Criteria). وتعرف هذه المحكات أحيانا بمؤشرات الأداء Performance Indicators.
- **مراجعة قواعد التقدير المتوفرة، أو المنشورة، أو التي صممها آخرون والمتعلقة بتقدير الأداء موضع البحث.** كذلك، يمكن مراجعة ماكتب الباحثون والمهتمون أو الخبراء عن الخصائص التي يجب أن يتصف بها هذا الأداء.
- **تجميع** مجموعة من أعمال الطلبة، ثم **تصنيفها** في مجموعات حسب مستوى نوعيتها: عالي، متوسط، ضعيف.
- **كتابة الخصائص والملامح** التي تجعل هذه الأعمال تختلف عن بعضها كالآتي:

المستوى	وصف مستوى الأداء
عالي	يمكن الاسترسال بقراءة النص، كما يمكن تتبع تدفق قطار الأفكار بمتعة فائقة، عبارات ممتعة، ومترابطة، ومثيرة. تم ابتداء الفقرات وختمها بأسلوب جذاب وذي معنى. تم الاستخدام المناسب للكلمات، تضمن النص عبارات متفاوتة بالطول، تم مراعاة قواعد الضبط النحوي بدرجة عالية من الدقة.
متوسط	الموضوع واسع، يمكن معرفة الاتجاه الذي يرنو إليه الكاتب. يمكن مواصلة القراءة، ولكن المعنى عموما غير واضح. يمكن أن تكتب بدايات الفقرات وخواتمها بأسلوب أفضل. الكلمات صحيحة ولكن الأسلوب غير مشوق. روعيت القواعد النحوية بدرجة متوسطة.
رديء	يحتوى النص على قائمة من المعلومات، لكنه يفتقر للمعنى.لا تختلف أهمية الموضوع عن أي شيء عادي.الموضوع ممل، ولا يشد القارىء إلى الاستمرار. لا يحوي النص موضوعا ذا أهمية. تساوي العبارات بالطول. وتمتع النص بعدد غير قليل من الأخطاء النحوية.

- **يتم اقتراح عدد من النقاط** التي يجب أن تعطى لكل مستوى من مستويات الأداء التي تم استنتاجها.
- **تجميع عينة من أعمال الطلبة (8-12) ومن ثم تصحيحها** في ضوء القاعدة التي تم تطويرها مبدئيا، ويتبع ذلك إجراء التعديلات المطلوبة لتطوير القاعدة على نحو أفضل.
- هذه الخطوات تنتهي بقاعدة **تقدير أداء كلية** Holistic Rubrics. أي أنها تنتهي بإعطاء علامة نهائية للطالب على الأداء ككل. في بعض الأحيان، قد يصعب على البعض تصنيف أداء أحد الطلاب بدقة كبيرة. فقد يكون هذا الطالب جيدا في كتابة مقدمة موضوعه وفي اختيار الكلمات المعبرة، في حين لم يكن موفقا في اختيارات كلمات أو عبارات بداية كل فقرة. هنا تظهر الحاجة إلى وجود قاعدة تقدير

تحليلية Analytical Rubrics. في هذه الحالة، لا بد من تطـوير أو استنتاج عدد من الأبعاد التي في ضوئها يمكن تقدير أداء الطالـب. وهي في حالة الكتابة مثلا: **الأفكار/ المحتوى، التنظـيم، التـأثير، اختيار الكلمات، طلاقة العبارات، العرض، الضوابط كما هو مبـين تاليا:**

عناصر الكتابة	ردىء (1)	متوسط (2)	عالي (3)
الأفكار / المحتوى Ideas Contents			
التنظيم Organizing			
التأثير Voice			
اختيار الكلمات Word Choice			
طلاقـة العبـارات Sentence Fluency			
العـرض Presentation			
الضوابط النحوية Conventions			

موجهات عامة لتطوير قواعد التقدير

- **من المفروض أن تنسجم محكات قاعدة التقـدير مـع متطلبـات المهمة والأهداف التي تم تحديدها مسبقا.**
- **من المفروض أن يتم صياغة محكات القاعدة بمثابة سلوكات قابلة للملاحظة:** فالمعلم لا يمكنه تقييم عمليات داخلية ما لم يتم ملاحظتها أو إظهارها. فتقييم الاستدلال مثلا، لا يتم إلا بتقييم ما يكتبه المتعلم أو يسلكه لفظيا.
- **من المفروض أن يتم كتابة قاعدة التقدير بلغة واضحة يستطيع أن يفهمها الطلبة:** إذ أن من فوائد استخدام قواعد التقدير أنـه يـزود الطلبة بوصف واضح لما هو متوقع قبل تقييم النشاط.
- **من المفروض أن يكون عدد النقاط المستخدمة في قاعدة التقـدير منطقية:** أي أن الأرقام التي ستعطى للقواعد التحليليـة أو الكليـة يجب أن تعكس القيمة الفعلية للنشاط.
- **يزداد مدى تقدير المهمة كلما صعبت مهمة الأداء أو كلما كانـت المهمة مفتوحة النهاية.**
- **المفروض أن يكون** عدد النقاط كافيا لتمييز الصفة. إذ عندما يكون العدد أقل من (4)، فهو غير كاف، بينما لا ينصح بأن تكون النقاط أكثر من (8). بالرغم من أن هذا العدد مقبولا في حالـة متـصل النمو (التطور) أو عندما يصعب التمييز بين المستويات.

- **في حالة أن يكون عدد النقاط = 4، فإن---**

المستوى (3) يشير إلى أن الأداء صحيحا وعاديا، ويحقق المطلوب.
المستوى (4) يشير إلى أن الأداء استثنائيا outstanding أو إبداعيا.
المستوى (2) ويشير إلى أن الأداء صحيحا لكنه غير مكتمل،
المستوى (1) ويشير إلى أن الأداء دون المستوى المقبول.

- **من المفروض أن يكون التفريق بين مستويات قاعدة التقدير واضحا:** بشكل عام، يفضل استخدام قواعد التقدير ذات العدد الأقل من المكونات على أن تكون الفروق والفواصل بين تلك المكونات واضحة ومميزة.
- **من المفروض أن تكون المحكات عادلة وخالية من التحيز:** أي أن تتم الصياغة بطريقة تتجنب أي تحيز تجاه الجنس أو العرق أو اللغة أو غيرها.

موجهات عامة لتطبيق تقييمات الأداء

1. **المفروض أن يتم إعداد المهمات الشفوية أو المكتوبة بدقة ووضوح وأن تصاغ بلغة يفهمها جميع الطلبة:** وهذا يستوجب مسبقا معرفة المستوى الدراسي للمتعلمين ومدى إتقانهم للغة التي كتبت فيها المهمة، كما أنه من المفروض أن يعطى الطلبة الفرصة الكافية للإستفسار عن معاني الكلمات التي تحويها المهمة.
2. **المفروض أن تتوفر كافة المواد والأدوات اللازمة للقيام بالمهمة كاملا:** وهذا يعتمد على نوع المهمة ومدى تعقيدها، فقد تتطلب بعض المهام الحصول على موارد ومراجع من المكتبة، أو أجهزة ومواد وأدوات من المختبر، أو برامج الحاسوب، والآلة الحاسبة بمستوى معين.

3. **المفروض أن يتم مناقشة قواعد التقدير مع المتعلمين قبيل الانتهاء من المهمة:** وهذا يسمح للمتعلمين أن يوائموا جهودهم بطريقة تعظم أداءهم . ويرى المعلمون أن إعطاء المتعلمين فكرة عن محكات التصحيح مسبقا، يجعلهم ذل جهد أكبر بهدف الحصول على أفضل مستوى أداء أفضل (على الرغم من ندرة حصول ذلك).

موجهات في تصحيح وتفسير نتائج تقييم الأداء

1. **من المفروض أن يكون المصححان المستقلان قادرين على تقدير درجات المتعلمين باتساق:** ويساعد في هذا أن تكون صياغة مكونات قاعدة التقدير بلغة واضحة وبمعنى دقيق.
2. **من المفروض أن يكون المصحح الواحد قادرا على إعطاء درجات متسقة على مدى فترات زمنية متعاقبة.** وهنا ربما تعمل بعض العوامل مثل معرفة المصحح المسبقة بالمتعلمين، ودرجة الحالة المزاجية للمصحح في وقت معين ربما تؤثر سلبا على صدق التقدير. وبشكل أدق، يطلب من المصحح الرجوع المستمر إلى قاعدة التقدير أثناء التصحيح.
3. **من المفروض أن تستخدم مجموعة الأعمال أو الأوراق (المثالية أو المرجعية: Anchor) في مساعدة المصححين في عملية تقدير الدرجات:** وهذه الأوراق التي يتم اختيارها كأمثلة للأداء على كل مستوى من مستويات قاعدة التقدير. والمفروض أن يتم رجوع المصححين لتلك الأوراق بشكل متكرر لضمان استقرار واتساق التصحيح على مدى اختلاف الوقت.

4. **يمكن أن تستخدم الأوراق أو الأعمال المرجعية بعد إزالة أسماء أصحابها للتوضيح للطلبة وأولياء أمورهم عن مستويات التصحيح المتضمنة في قواعد التقدير.** هذا الإجراء، يمكن أن يساعد في كشف الغموض الذي ربما يظهر في مستويات قاعدة التقدير.

5. **من المفروض أن يكون الاتساق بين الدرجة النهائية وقواعد التقدير واضحا وفوريا:** فعند استخدام قاعدة التقدير التحليلي، فإن الدرجة التي يتم تقديرها على كل مستوى يجب أن تظهر وتوثق.

6. **من المفروض أن تستخدم النتائج على مهام الأداء في تحسين التدريس وطرق التقييم.** إذ ما ذا استفاد المعلم من استجابات الطلبة؟ وكيف يمكن استخدام ذلك في تحسين التدريس الصفي مستقبلا؟ ماذا يمكن للمعلم أن يعمله في تطوير قواعد التقدير التي يستخدمها في تقييمه لطلابه؟

فيما يلي ثلاثة قواعد تقدير يتم الاطلاع على كل منها للفائدة:

نموذج قاعدة تقدير لمهارة "تحديد الوقت"

A Rubrics for a skill of "Telling Time"

تحديد الوقت Telling Time	يحتاج مساعدة Needs Help (1)	يتطور Developing (2)	يصل حد الإنجاز Achieving (3)
يحدد الوقت بالساعة ونصف الساعة	يمزج الطفل بين ذراع الساعة وذراع الدقيقة	يحدد ذراع الساعة ويجهل تحديد موقع ذراع الدقيقة	يحدد الوقت تحديدا صحيحا
يحد الوقت بدلالة وحدات الخمس دقائق	يحتاج إلى مساعدة في تحديد الوقت	يمزج الطفل بين ذراع الساعة وذراع الدقيقة	يحدد الوقت تحديدا صحيحا
يبين الوقت باستخدام ساعة أحد الطلاب	لا يستطيع وضع أي من الذراعين في موضعه	يمزج الطفل بين ذراع الساعة وذراع الدقيقة	يضع الذراعين في موضعهما الصحيح
يكتب الأرقام على نموذج أصم للساعة blank	يضع الأرقام في مواضع عشوائية	يضع معظم الأرقام في مواقعها	يضع جميع الأرقام في مواقعها الصحيحة
يرسم الوقت على صورة تبين ساعة	لم يضع أي من الذراعين في الموقع الصحيح	يمزج الطفل بين ذراع الساعة وذراع الدقيقة	يضع الذراعين في موضعهما الصحيح

نموذج قاعدة تقدير لمهارة "المثابرة" Perseverance

هل أنت إنسان مثابر Do You Show Perseverance?	نادرا Hardly Ever	أحيانا Some of the Time	غالب الأحيان Very Often
أحاول تحقيق الأمر مرات حتى أحصل عليه	1	2	3
أنا لا أتردد في التقدم حتى ولو أصبح التقدم صعبا	1	2	3
أنا أحاول أكثر من مسلك إذا احتجت إلى ذلك	1	2	3
أعود إلى قراءة التعليمات عندما أواجه أي عقبة	1	2	3
أعرف الحصول على معلومات إذا احتجت إليها	1	2	3
أتفحص عملي لتحقيق الدقة	1	2	3
أحاول تجريب البدائل المتاحة لحل المشكلات	1	2	3

نموذج مقترح لقاعدة تقدير "مهارة قراءة وتفسير الخرائط"

مؤشرات النجاح	مكتشف عالمي	رحالة حول العالم	زائر متجول	قابع في بيته
يستطيع الطالب تفسير رموز الخريطة	4	3	2	1
يستطيع الطالب استخدام وتفسير دليل الخريطة	4	3	2	1
يستطيع الطالب تحديد المواقع	4	3	2	1
يستطيع الطالب استخدام مقياس الرسم لتحديد المسافات	4	3	2	1
يستطيع الطالب وصف المواقع النسبية	4	3	2	1
يستطيع الطالب مقارنة الخرائط وعمل التفسيرات	4	3	2	1
يستطيع الطالب تفسير المعلومات واستنتاجات صحيحة	4	3	2	1

أنظمة الدرجات والعلامات المتعدد

يلخص الجدول التالي إيجابيات وسلبيات الطرق المختلفة التي تستخدم في التعبير عن مدى تقدم الطلبة. وتختلف هذه الطرق باختلاف المنطقة التعليمية وباختلاف الجهة المستفيدة أو المستخدمة لتلك الدرجات. وقد يحدث هناك تداخل أو تناقض حسن استخدام عدة طرق معا، فقد يلجأ إداريو المدارس إلى إعداد تقارير دقيقة عن مستوى تقدم الطالب تستخدم في تكوين سجلات يسهل حفظها والرجوع إليها، في الوقت الذي يفضل أولياء الأمور الحصول على معلومات عن أبنائهم تتسم بالشرح المفصل عن المحتوى الذي تم تدريسه للطالب وما المستويات أو المعايير التي توصل إليها الطالب وما موقعه بالنسبة للمجموعات المعيارية.

إيجابيات وسلبيات الطرائق المستخدمة في إعداد تقارير تقييم الطلبة

نوع الطريقة	الرمز المستخدم	الإيجابيات	السلبيات
الدرجات بالحروف	أ، ب، ج.....الخ ويمكن إضافة + أو – كما في (أ–) أو (ب+) مثلا.	أ) سهلة الإدارة والاستخدام. ب) يعتقد بأنها سهلة التفسير. ج) تلخص بدقة الأداء ككل.	أ) يتنوع معنى الدرجة بشكل واسع حسب المادة والمدرس، ومن مدرسة إلى أخرى. ب) لا تصف نواحي القوة أو تحدد مواطن الضعف. ج) يشعر طلبة رياض الأطفال والابتدائية عند استخدامها معهم بنوع من الخيبة.
أرقاما أو نسبا مئوية	أرقام صحيحة (3، 4، 5) أو نسب مئوية مثل: 99%، 98 % .	أ) سهلة الإدارة والاستخدام. ب) يعتقد بأنها سهلة التفسير. ج) متصلة أكثر من الحروف. د) يمكن استخدامها جنبا إلى جنب مع الحروف.	أ) يتنوع معنى الدرجة بشكل واسع حسب المادة والمدرس، ومن مدرسة إلى أخرى. ب) لا تصف نواحي القوة أو تحدد مواطن الضعف. ج) يشعر طلبة رياض الأطفال والابتدائية عند استخدامها معهم بنوع من الخيبة. د) لا يبدو معنى كل منها واضحاً مباشرة، ما لم يرافقه تفسير مناسب أو توضيح.
درجات تتألف من فئتين	ناجح، راسب مرضي، غير مرضي Satisfactory, Notsatisfactory	أ) أقل إيذاء للطلبة الصغار. ب) يشجع الطلبة الكبار على التسجيل في مواد	أ) أقل ثباتا من النظم المتصلة أو تلك الأكثر من فئتين. ب) لا يوصل معلومات

		غير مهمة خوفا من انخفاض معدلاتهم التراكمية.	كافية حول أداء الطلبة للآخرين للحكم على مدى التقدم والتحسن.
قوائم فحص وسلالم تقدير	إشارة (x) أمام الأهداف التي تحققت أو تقديرات رقمية لدرجة التحقق.	أ) يعطي تفصيلا لمستوى تحصيل الطالب. ب) يمكن استخدامها بجانب التقدير بالحروف.	أ) يثقل فهمه على بعض أولياء الأمور نظرا لكثرة التفاصيل. ب) يصعب حفظه في السجلات.
تقرير قصصي (سردي)	لا يوجد رموز محددة، وقد يستخدم أي من التقديرات أعلاه، ولا تنتهي إلى درجات.	أ) يتيح الفرصة للمعلم أن يصف مدى تقدم الطالب أكاديمياً. ب) يعرض تقدم الطالب وفقاً لمعايير أو مؤشرات للإنجاز أو الأهداف التعلمية. ج) يتيح الفرصة لفتح حوار أو اتباع وسائل اتصال معينة مع أولياء الأمور والطلبة.	أ) يستغرق وقت طويل. ب) يحتاج إلى مهارة في الكتابة والتواصل الفعال من قبل المدرس. ج) يؤدي تبسيط اللغة كي يقبلها أولياء الأمور إلى احتمال فقد بعض المعاني المهمة. د) يلجأ بعض أولياء الأمور إلى إهمال التقرير أو عدم فهمه نظرا لافتقارهم قدرات لغوية معينة. ه) قد لا يستجيب بعض من أولياء الأمور. و) غالبا ما يتم تعديله بإضافة بعض قوائم الفحص أو مؤشرات وتعليقات مختصرة من قبل المدرس.
لقاءات الطلبة والمدرسين	لا يوجد رموز و لكن يمكن مناقشة أي الرموز أعلاه.	أ) يتيح الفرصة لمناقشة مدى التقدم في الأداء بشكل شخصي. ب) يمكن أن يكون عملية مستمرة تتكامل مع	أ) يحتاج المدرس إلى مهارة تقديم الملاحظات بأنواعها إيجابية وسلبية. ب) يستغرق وقتاً طويلاً. ج) يشكل تهديدا لبعض

		التدريس.	الطلبة. د) لا يقدم سجلات مختصرة وملخصه للمدرسة.
لقاءات أولياء الأمور و المدرسين	لا يوجد رموز و يمكن مناقشة أي من الرموز أعلاه.	أ)يتيح الفرصة أمام أولياء الأمور والمعلمين لمناقشة الاهتمامات وتوضيح سوء الفهم. ب) يمكن المدرسين من عرض نماذج من أعمال الطلبة وتفسير أسس تقييمها والحكم على نوعيتها. ج) يحسن علاقة المدرسة بأهل المتعلم.	أ) يستغرق الوقت الكثير. ب) يتطلب تحضيراً مسبقاً من المدرس. ج) قد يبعث القلق لدى بعض المدرسين وأولياء الأمور. د) لا يعد وسيلة فاعلة في توصيل كميات كبيرة من المعلومات. ه) قد يكون من غير السهل حضور بعض أولياء الأمور.
مراسلة أولياء الأمور (التواصل)	لا يوجد رموز ويمكن استخدام واحد أو أكثر من أعلاه.	تساهم في تدعيم الفائدة المرجوة من التقارير الأخرى.	أ) الرسائل القصيرة لا تصف أو تعبر عن مدى تقدم الطالب بفاعلية. ب) تتطلب مهارة فائقة وكثير من وقت المدرس.

الدرجات مالها وما عليها

يختلف الأشخاص في كيفية الاستفادة من الدرجات التي يحصل عليها المفحوصون. فقد تختلف استفادة المعلم من الدرجات عن استفادة كل من المرشد النفسي أو مدير المدرسة أو ولي الأمر، أو حتى الجهات الأكاديمية كما في الجدول التالي:

استخدامات الدرجات حسب الجهة المعنية (مستخدمو الدرجات)

استخدام الدرجات	مستخدمو الدرجات Stakeholders						
	الطالب	ولي الأمر	المعلم	المرشد	إدارة المدرسة	مؤسسات التعليم	أصحاب الأعمال
للتأكد من مستوى التحصيل الصفي.	X		X				
لتوثيق مدى التقدم التربوي.	X	X	X	X	X	X	X
للتحقق من مدى إكمال متطلبات المادة الدراسية.	X	X	X	X	X	X	X
للتوجيه بمنح المكافأة أو توجيه العقوبة.	X	X					
لجذب انتباه المعلم والإهتمام الاجتماعي.	X						
لتحديد التسكين التربوي الجديد المطلوب.		X	X	X	X		
للحكم على كفاية المعلم وعدالته.		X			X		
للإشارة إلى مشكلات المدرسة المتعلقة بالطالب.		X	X	X	X		
لدعم برامج التوجيه المهني والوظيفي.	X	X		X			
للحد من/أو وقف مشاركة الطالب في النشاطات اللامنهجية.		X	X		X		
لأغراض الترقية والنقل أو الإعادة.			X		X		
لمنح شهادات التخرج.					X		
لتحديد ما إذا كان الطالب يحتاج لدراسة متطلبات سابقة ما.			X	X	X		
لأغراض الإلتحاق بالدراسات العليا.						X	
للتأكد من مدى امتلاك الفرد المهارات الأساسية اللازمة لمهنة معينة.							X

ويتضح من البيانات الواردة في هذا الجدول أن الاستخدام رقم (2) المتعلقة بتوثيق مدى التقدم التربوي والاستخدام رقم (3) المتعلق بإكمال المتطلبات السابقة هما الأكثر استخداما من قبل كافة المعنيين ومستخدمي الدرجات معلمين ومرشدين وأولياء أمور وحتى مؤسسات التعليم العالي. في الوقت الذي تتفاوت فيه كثافة الاستخدامات الأخرى للدرجات من مستفيد إلى آخر.

نقد الدرجات والعلامات Criticisms of Grades and Marks

دأب المربون على توجيه العديد من الانتقادات للدرجات على مدى من السنين. ويبدو أنه من الطبيعي أن نعي تلك الانتقادات لنتمكن من شرح مبررات سياسة تقدير الدرجات لأولياء الأمور والمربين الآخرين وحتى المهتمين. وفيما يلي تصنيف لتلك الانتقادات الأربع متبوعا بالتعليقات المناسبة:

I. ليس للدرجات معنى أساسيا وثابتا Grades are Essentially Meaningless

وينبع هذا النقد من خلال استعراض البنود التالية:

1. هناك تنوع كبير بين المدارس أو المدرسين في تقدير الدرجات.
2. ينقص العديد من المدارس سياسات منهجية في تقدير الدرجات.
3. لا يستطيع مجرد رمز، التعبير عن تفاصيل معقدة تتعلق بمستوى أداء الفرد.
4. عادة ما يكون المدرسون غير مبالين أو جادين في تقدير الدرجات.
5. يساء أحيانا استخدام الدرجات، فقد تستخدم كإجراء عقابي، أو فرضا للالتزام بالأنظمة والسلوك الهادىء أكثر من كونها تعبيرا عن مستوى الأداء.

ثمة بعض الحقيقة في هذا النقد. إذ يمكن استخدام البنود 1، 2، 4، 5 في دعم وتأييد الحاجة إلى تحسين وتقوية طرق وممارسات تقدير الدرجات، وليس التخلص منها أو إلغائها. أما البند رقم (3) فإنه ينسب للدرجات أكثر مما يمكن أن توصله أو تبلغه. وليس المقصود من الدرجات الرمزية والحروف المختصرة أن تحل محل التفاصيل التي تشرح تحصيل الطالب. إذ أن الدرجات تقارير مختصرة حول مستوى الأداء العام للطالب.

II. ليس للدرجات أهمية تربوية Grades are Educationally Unimportant

ويمكن تلخيص هذا الادعاء من خلال البنود التالية:

1. ليست الدرجات إلا مجرد رموز.
2. النتائج الهامة غير ملموسة ولا يمكن تقييمها بالدرجات.
3. درجات المعلم أقل أهمية للطالب من تقييمه الذاتي.
4. لا تتنبأ الدرجات بمستوى التحصيل في المراحل القادمة بدقة.
5. يحتاج البرنامج التربوي إلى التقييم المستمر مثل أو أكثر من حاجة الطلبة له.

يعد هذا النوع من النقد مزيج من عبارات حول القيمة من ناحية، والمقارنات غير الضرورية من ناحية ثانية. فمما لا شك فيه، فإن الدرجات رموز، ولكن هذا لا يعني أن الرموز ليست هامة. فبعض الأشخاص يعملون للوصول إلى مكافأة محسوسة ومادية، والبعض الآخر يعمل من أجل الحصول على مكافأة رمزية ومعنوية. فإذا نظرنا إلى النتائج غير المحسوسة على أنها غير هامة، فإن ذلك يعني أن التعليم ليس له أية قيمة ولا يمكن ملاحظته أو تقييمه. ويبدو هذا أمر يتعذر الدفاع عنه. وعليه، يتعين علينا الأخذ بتقييم المدرس والدرجات التي يرصدها، إضافة إلى تقييم الطالب لنفسه.

فالتقييم الذي يحدده المدرس يساعد الطالب على تقييم نفسه بشكل واقعي. وما الدرجات إلا مؤشرات تنبؤ حول تحصيل الطالب الأكاديمي، كما أنها تتبيء عن بعض أنواع الإنجازات خارج الفصل الدراسي.

من زاوية أخرى، ليس من المتوقع أن تكون الدرجات كاملة أو شبه كاملة في التوقعات. فهناك عوامل تلعب دوراً هاماً مثل الفرصة، والجهد، ونوعية التعليم والحظ. علاوة على أنه ينبغي أن تعكس الدرجات المواضيع التي درسها الطالب، فإذا لم تتوافق أهداف المنهاج مع المهارات والقدرات والميول الضرورية لنجاح العمل، فإننا نتوقع أن يكون دور الدرجات ضعيفاً في التنبؤ، حتى لو كانت تعكس ما تعلمه الطالب.

III. ليست للدرجات ضرورة Grades are Unnecessary

وقد انبثق هذا النقد من خلال استعراض المعطيات التالية:

1. الدرجات لا تلعب دورا كونها مثيرات motivators فعالة للتحصيل الفعلي في التعليم.
2. عندما يتعلم الطالب حتى مستوى الإتقان، لا تظهر فروق بين مستويات أداء الطلبة لتقييمها.
3. ما زالت الدرجات أسلوب شائع في تقييم أداء الطلبة، وذلك بسبب تعلق المدرسين بالممارسات التقليدية وابتعادهم النسبي عن التغيير.

توضح النقاط السابقة أن الدرجات لا يمكن التخلص منها وإلغائها تماماً، هذا على الرغم من كونها غير ضرورية بالكامل لأغراض التقييم. ثمة حاجة إلى تقرير مختصر لأغراض الإرشاد والمرجعية، وتحديد المسؤولية لكل من الطلبة تجاه المدرسة، ومسؤولية المدرسة تجاه أولياء الأمور. فالبيانات تحتاج إلى تلخيص وتنظيم من أجل أن يتم استيعابها بدرجة كبيرة وهكذا.

فبالرغم من عدم كمال الدرجات، فإنها تعد تلخيصاً وسجلا للحفظ. فقد يحتاج أولياء الأمور معرفة مدى تحصيل أبنائهم في الرياضيات على سبيل

المثال. أو ربما يود هؤلاء معرفة كفاءة الأبناء في مواضيع محـــددة فـــي الرياضيات. عند ذلك قد لا تكون التقارير المفصلة مفهومة. فعلـــى ســـبيل المثال، قد يرغب طالب في المرحلة الثانوية مثلا معرفة هل يسجل لمـــادة حساب التفاضل والتكامل العادي أو المتقدم. فيقوم مرشـــده بـــالنظر إلـــى درجاته في المادة التي تسبق حساب التفاضل والتكامل (المتطلب الـــسابق)، ويلاحظ أن الطالب لم يتعلم أية مفاهيم حول التفاضـــل والتكامـــل خـــلال الأسابيع التسعة الأخيرة، عندما يدله المرشد نحو الصواب. وعندما يرغب طالب في معرفة مدى تقدمه في التحصيل، يتعين على المـــدرس أن يقـــدم قراراً وحكماً ودرجات، وليس مجرد وصف عن الصف الدراســـي الـــذي وصل إليه.

IV. تشكل الدرجات مصدرا للألم Grades are Harmful وذلك من خلال البنود التالية:

1. تثبط الدرجات المتدنية الطلبة الأقل قدرة من بذلهم الجهـــد الـــلازم للتعلم.
2. تؤدي الدرجات إلى جعل الفشل أمرا لا مفر منه لدى بعض الطلبة.
3. يعاقب أولياء الأمور أحيانا أبناءهم إذا حصلوا على درجات متدنية، ويكافئون أصحاب الدرجات المرتفعة بشكل غير مناسب.
4. تشكل الدرجات معايير عالمية لكافة الطلبة رغم اتـــساع الفـــروق الفردية بينهم.
5. تركز الدرجات على أهداف عامة لجميع الطلبة ولا تشجع الفرديـــة في التعلم.
6. تكافىء الدرجات الامتثال وتعاقب الإبداع والابتكار.
7. تشجع الدرجات المنافسة أكثر من تشجيعها التعاون وروح الفريق.

8. ربما يؤدي ضغط الحصول على الدرجات العالية لجوء بعض الطلبة إلى الغش.
9. تتواءم الدرجات مع التعلم المرتبط بالموضوع أكثر من التعلم المرتبط بالمتعلم.

يستخدم بعض المدرسين الدرجات لمعاقبة الطلبة والثأر منهم، وليس أداة تربوية. كما أن بعض أولياء الأمور يركزون كثيراً على أهمية الدرجات، ويضغطون على أبنائهم ليحصلوا على درجات عالية بأي ثمن. وجدير بالقول بأنه لا تتوفر بيانات حول نسبة عدد المدرسين الذين يعاقبون الطلبة، أو حجم أولياء الأمور الذين يضغطون على أبنائهم.

ومن الأمور التي تستدعي الاهتمام والتأثير المثبط للدرجات المتدنية، قام كرونباخ بتحليل تأثير الدرجات واقتراح بعض الطرق التي يمكن للمدرسين اتباعها لتحسين التقييم للحد من إحباط الطلبة. وبالرغم من أن الدرجات قد تكشف فقط فشلاً في التعلم، فإنها لا تلام بالكامل على هذا الفشل، فإنه من الأهمية دراسة كيفية تأثير تلك الدرجات على الطلاب (Nitko, 2005).

الاتجاهات نحو الدرجات والعلامات

في نهاية هذا الفصل أو حتى في بدايته، كان من المهم التعرف على شعور الطالب نحو عملية تقدير الدرجات والعلامات. وعليه، يتوقع منك أيها الطالب قراءة العبارات التالية في الجدول التالي، ومن ثم ضع إشارة تحت (موافق) إذا كنت توافق، أو إشارة تحت (غير موافق) إذا كنت غير موافق، وإشارة تحت (غير متأكد) إذا كنت غير متأكد. قارن إجاباتك مع إجابات زملائك أو مدرسيك. وحاول أن تحافظ على هذه الاتجاهات بداخلك

وأنت تدرس هذا الفصل، ثم فكر في كيف يمكن استخدام المفاهيم أثناء تدريسك مستقبلا.

ما شعورك نحو الدرجات والعلامات؟

الرقم	العبارة	موافق	غير متأكد	غير موافق
1	يفضل استخدام أكثر من ثلاث فئات من العلامات في المجال الأكاديمي.			
2	هناك أسباب مقنعة حول كون متوسط درجات بعض المدرسين أعلى من غيرها وبثبات.			
3	إذا أخفق طالب في مساق (مقرر أو مادة دراسية)، ثم اجتازه. فهل يجب أن تبقى الدرجة الأولى في سجل الطالب.			
4	يجب أن يأخذ طلبة الجامعة دورا في تقويم أعمالهم وتصحيح منجزاتهم.			
5	يجب أن تركز العلامات الأكاديمية على مستوى التحصيل أكثر من تركيزها على مقدار التقدم في مستوى الأداء.			
6	إذا تم استخدام العلامات، فالمفروض أن تكون بالأرقام وليس بالحروف.			
7	علامات الطلبة في المجال الأكاديمي يجب أن تحدد على أسس التحصيل الأكاديمي وليس على أساس الحضور والمواظبة، والمواطنة، والنشاط، والاتجاهات.			
8	يجب على الجهات (مدارس أو كليات) التي تستخدم العلامات أن تعتمد وتعلن سياستها المؤسسية عن			

	التقييم.			
9	عندما تعطى العلامات للطلبة في المجال الأكاديمي، فيجب عرضها كقياسات وليس تقويما.			
10	يجب أن تعطى نفس نسبة الدرجات العالية للصفوف الطلبة بطيئي التعلم كما في صفوف سريعي التعلم.			
11	يستخدم معظم المعلمين عددا قليلا من الأساليب الإحصائية في تقويم طلابهم.			
12	بغياب سياسة العلامات المؤسسية، فإنه لا يعقل استخدام العلامات في الحكم على أهلية الطالب للالتحاق بالبرامج الأكاديمية.			
13	يشير معظم النقد القائم للعلامات إلى العلامات فقط وليس إلى تدريبات تصحيح محددة.			
14	تعتبر المعايير المطلقة أكثر تفضيلا من المعايير النسبية في تقويم الطلبة وإعطاءهم الدرجات			
15	بغياب سياسة العلامات المؤسسية، فإنه لا يعقل استخدام العلامات في الحكم على أهلية الطالب للاشتراك في الأنشطة الرياضية وغير الصفية			
16	ناجح/راسب أفضل من العلامات بثلاث فئات أو أكثر في المستوى الأكاديمي			
17	السماح للطلبة للتعاقد حول علاماتهم أكثر تفضيلا من استخدام العلامات على أساس نسبي			
18	يعتمد غالبية المعلمين على أحكامهم الذاتية في تقويم طلبتهم			
19	بعض أنظمة التقييم (الأرقم أو الحروف) أساسي في الممارسة التربوية الجيدة			

20	تخلق الممارسات العادية في إعطاء العلامات منافسة غير مقبولة بين الطلبة				
21	يجب أن تعتمد علامات الطلبة في نهاية المادة الدراسية على مستوى التحصيل نهاية المادة الدراسية وليس على معدل تحصيل كل منهم على مدى الفصل الدراسي.				
22	يجب استخدام عوامل مثل الاتجاهات والاهتمامات في تحديد علامات الطلبة				
23	في حالة وجود شيء ما ، فإنه موجود بمقدار، وعليه فإنه يمكن قياسه.				
24	يجب أن يحاول المعلمون تقويم الطلبة في مجالات معينة مثل الاهتمامات والاتجاهات والدافعية				

المهارات الأساسية الضرورية عند الأداء على أداة التقييم

لا بد للطلبة من تعلم المهارات التالية أثناء الأداء على أنواع التقييم وخاصة الامتحان(Ebel and Frisbie, 1991):

- الانتباه إلى التعليمات شفهية وتحريرية وإدراك التبعات المترتبة على عدم متابعتها.
- معرفة كيف سيتم تصحيح التقييم، وكيف سيتم تقدير أهمية المهام بالنسبة لكامل التقييم، وكيف سيتم خصم النقاط مقابل الإجابات الخاطئة كما في الأخطاء النحوية والكتابية.
- كتابة الاستجابات واختيار الإجابات (بوضع إشارة أو دائرة حول رمزها) بدقة وترتيب، لتفادي خسارة بعض الدرجات بسبب عدم القيام باختيار الإجابة المطلوبة بخط واضح أو بتأشير غير دقيق.

- ضرورة الدراسة والمذاكرة بشكل دوري ومنتظم قبيل موعد التقييم مخافة التعب والإرهاق.
- استثمار وقت التقييم بحكمة، لضمان أن تتم الإجابة على كافة المهام خلال الوقت المخصص.
- استخدام المعرفة الجزئية والتخمين عند الإجابة بشكل مناسب.
- ترتيب وتنظيم الإجابات على الفقرات الصياغية قبل البدء بالكتابة: وضرورة تحديد الوقت الملائم لكل جزء منها.
- تفحص العلامات التي تعطى لكل جزء على أوراق منفصلة، لتجنب أن لا تتم خسارة النقاط التي سبق تخصيصها لجزء تم حذفه.
- مراجعة إجابات المهام وتعديل الإجابات إذا كان ذلك يحقق الإجابة الأفضل.

حكمة الإختبار Test Wiseness

تعرف حكمة الإختبار بأنها مقدرة المفحوص على (1) استخدام استراتيجيات التقييم، و (2) الضعف في صياغة الفقرات ووجود المنبهات، و(3) خبرة المرور بأكثر من تقييم في تحسين درجته أكثر مما يستحق. فعندما تقوم (أيها المعلم) بصياغة اختبار صفي مثلا، كن على وعي من أن عددا من الطلاب يستفيد من أخطائك في صياغة فقرات الإختبار ويتم كسب عدد من الدرجات دون تمكنهم من المادة الدراسية. ويبين الجدول التالي النقاط العامة التي تتكون منها استراتيجيات حكمة الإختبار.

المفروض أن يتم بناء تقييم يتمتع بجودة عالية لا يسمح لمن لم يصل إلى المستوى المطلوب من تحقيق درجات لا تصف مستوى الطالب بصدق. كما أن المطلوب تعليم الطلبة المهارات الواردة في الجدول المذكور لتقليل التباين بينهم وبين من يستثمر مهارات حكمة الإختبار من

زملائهم. ومن المعروف أن حكمة الإختبار تتطور لدى الفرد مع النضج، وبتطور الطالب في الدراسة، والخبرة (Sarnacki, 1970; Slakter, Kochler, Hamton, 1970).

مهارات حكمة الإختبار

قبيل القراءة المستفيضة للتقييم، ابدأ بالأسئلة القصيرة، وحاول الإجابة على كل واحدة منها. وحتى لو لم تكن متأكدا تماما من الإجابة، فهناك الإجابة الصحيحة أو الأقرب إلى الصحة لكل فقرة. وفيما يلي عدد من الفقرات المشتقة بقليل من التعديل من دايموند وإيفانز (Diamond and Evans, 1972):

1. Harold Stone's book, *The Last Friendship*, is an example of an:
   a. Political satire
   b. Autobiography
   c. Science fiction
   d. Biography

2. عدد سكان مدينة فرانك هو أكثر من ----:

a. 50 ألفا

b. 60 ألفا

c. 70 ألفا

d. 80 ألفا

هاتان الفقرتان ضعيفتان ولكن يمكن إجابة أي منهما إجابة صحيحة بسبب المنبهات الموجودة في متن كل منها. فالفقرة الأولى تكون إجابتها رقم (b) نظرا لوجود مؤشر قواعدي على ذلك، إذ أن كلمة "an" لا تأتي إلا قبل حرف من حروف العلة، وهذا موجود في البديل (b). أما الفقرة الثانية، فبدائلها تبدو متداخلة لوجود كلمة "أكثر من" في المتن.

مبادىء (استراتيجيات) حكمة الإختبار

الإستراتيجية	أ. عناصر مستقلة عن باني الإختبار أو أهدافه
استخدام الوقت	1. ابدأ الإجابة بسرعة ممكنة مع المحافظة على الدقة والصحة في الإجابة. 2. ضع خطة تقدم في الإجابة على التقييم (الإختبار). 3. يمكن أن تحذف أو تخمن في إجابة الفقرات التي لا يمكن إجابتها بسرعة. 4. حدد الفقرات التي تم حذفها أو التي تحتاج اهتماما زائدا. 5. استخدم الوقت المتبقي بعد الانتهاء من الإختبار في إعادة التأكد من الإجابات.
تجنب الأخطاء	1. الانتباه أكثر لتعليمات التقييم (الإختبار) وتحديد المطلوب بدقة من المهمة. 2. الانتباه أكثر لفقرات التقييم وتحديد طبيعة السؤال. 3. الاستيضاح من الفاحص إذا كانت التعليمات تسمح بذلك. 4. تفحص كافة الإجابات.
التخمين	1. يمكن التخمين دائما إذا كان الجواب الصحيح فقط هو الذي يحسب. 2. يمكن التخمين دائما إذا كان التصحيح لأثر التخمين أقل تطرفا من المعادلة التي تعطي صفرا للتخمين العشوائي.
الإستدلال الإستنباطي	1. استبعد البدائل التي تبدو خاطئة، ثم اختر الجواب من البدائل الباقية. 2. استبعد أو اختر كلا البديلين الذي يصحح بعضها الآخر. 3. استبعد كلا البديلين أو اختر واحدة منها فقط إذا كان أحدهما يبين خطأ الآخر. 4. استثمر المعلومات المتضمنة في بعض الفقرات في الإجابة على فقرة.
ب. عناصر تعتمد على باني الإختبار ومطوره (مدرس المادة)	

ا ل م غ ز ى	1. فسر وأجب على الأسئلة في ضوء تركيز باني الإختبار أو في ضوء أهداف الإختبار. 2. أجب على الفقرات بالطريقة التي يفضلها باني الإختبار. 3. اعتمد مستوى معين من التفصيل في الإجابة كما يتوقع (وتتعلق هذه النقاط بالفقرات الصياغية عموما)
استخدام المنبهات	1. استخدم المنبهات الممكنة في الإجابة على الفقرات: أ) فقد يبدو الجواب الصحيح أطول أو أقصر من البدائل غير الصحيحة. ب) قد يتم تسليط الضوء على الجواب الصحيح أو قد يصاغ بأسلوب مميز. ج) قد يكون الجواب الصحيح موضوعا بموقع متميز (في وسط البدائل مثلا). د) قد يكون الجواب الصحيح غير منسجم قواعديا مع البدائل الأخرى. 2. الحرص على تضمين المعلومات التفصيلية المتعلقة بالإجابة على الفقرة. 3. استخدم ما يتشابه من المتن في البدائل.

بيانات حول تغيير الإجابات

هل ينتفع الطلبة إذا قاموا بتغيير إجاباتهم حال نقلها إلى ورقة الإجابة؟ بصرف النظر عن هذا الرأي الشائع، فإن هذا ربما يعود إلى تغيير الإجابات إذا كان تغيير الفقرة يعتمد على إعادة التفكير بالفقرة. وهناك عدد من نتائج البحوث ذات العلاقة:

- يعتقد معظم التربويين والممتحنين أن ذلك ليس بسبب تغيير الإجابة.
- معظم الطلبة يغيرون إجاباتهم لما نسبته 4 % من الفقرات.
- أشارت بعض نتائج البحوث إلى أن ثلثي الإجابات تغيرت إلى الصواب.
- يقل مدى الاستفادة من تغيير الإجابات في حالة الفقرات الصعبة.

- يستفيد الطلبة ذو الدرجات المنخفضة أقل من ذوي الدرجات المرتفعة حال تغيير الإجابات.

قلق الإختبار Test Anxiety

على مدى الخبرات التربوية التي يمر بها الطلبة، هناك عدد من الطلبة الذين يجلسون أمام مدرسيهم وزملائهم في الصف للدراسة وإجابة الأسئلة خلال الدروس والجلوس للتقييم، ويختلف الطلبة في ذلك كثيرا. فمنهم من لا يبالي كثيرا. ومنهم من يكون متحفزا جدا لذلك، وبعض من هؤلاء المتحفزين تعمل التقييمات وزيادة الضغط النفسي عليهم (وهو ما يعرف بقلق الإختبار). فبعض من هؤلاء الطلبة يعتقد أن التقييم هو بمثابة تحد. وبعضهم يعتقد أن التقييم نوع من الخطر. وهؤلاء الطلبة ربما أنهم غير قادرين للقيام بأداء المهمة كما يجب، ولم يسبق أنهم تعلموا كيفية القيام بالمهمة.

أما الطلبة الذين يعتقدون بأن التقييم تحد، فإنهم يرون أن التقييم متعلق بالمهمة. وهؤلاء يتركز انتباههم وتفكيرهم على ضرورة إكمال المهمات وتقليل الضغوط المتعلقة بها. في الوقت الذي يعتقد الطلبة الذين يرون أن التقييم خطر بأن التقييم غير مرتبط بالمهمة. إنهم يركزون على ما سيحدث لهم في حال الفشل، ويفضلون الهرب من الموقف بالسرعة الممكنة.

الانفعالية والإهتمام Emotionality and Worry

يعتقد بشكل عام أن قلق الإختبار يتألف من مكونين هما الانفعالية Emotionality والاهتمام Worry. وتعرف الانفعالية بأنها الإثارة الجسدية التلقائية (الرجفة) المرتبطة مع قلق الإختبار. بينما يرتبط الاهتمام الجانب المعرفي للقلق: وهو يمثل مجمل الاعتقاد بالتبعات السلبية للفشل، وكيف سيكون أداء الفرد مقارنة بالآخرين.

وقد أشارت الدراسات إلى أن ردود الفعل الانفعالية تجاه التقييم الاهتمام والذي بدوره يؤدي إلى الانتهاء بأداء منخفض. وقد يعرف الاهتمام بأنه التفسير المعرفي. بمعنى أن الطالب المضطرب تجاه الأداء الضعيف للمهمة يجعل الطالب بمستوى عال من القلق بدلا من التركيز والانتباه على إنجاز المهمة المطلوبة. وكلما اتضح ضعف أدائه على التقييم.

العوامل التي تسبب قلق الإختبار

أمكن تصنيف ردود فعل الطالب تجاه التقييم في أربع عوامل مترابطة هي: التوتر، والاضطراب، والتفكير غير المرتبط بالإختبار، وردود الفعل الجسمية (Sarason, 1984). وجدير بالذكر أن هذه الأبعاد الأربع تبدو مناسبة لأفراد من جنسيات مختلفة. ويبين الجدول التالي نموذجا من الفقرات التي تقيم تلك الأبعاد. حيث تم وصف قلق الإختبار بوضوح أكثر مما تم في حالة الانفعالية والاضطراب. فقد تم التعامل مع الانفعالية على أنها تتكون من التفاعلات الجسمية والاضطراب، بينما كان التفكير غير المرتبط بالإختبار جزءا من قلق الإختبار جنبا إلى جنب مع الاضطراب.

عينة من فقرات أبعاد ساراسون لقلق الإختبار

التوتر Tension	• أشعر محزونا ومرتبكا قبيل الإختبار • أشعر بالنرفزة قبل الإختبار • أجد نفسي قلقا يوم الإختبار
الاهتمام Worry	• قبل أخذ الإختبار، أكون قلقا خوفا من الفشل • خلال الإختبار، أكون متعجبا عن الآخرين وكيف سيكون أدائهم. • قبل الإختبار، أشعر بمشكلة بما يمكن أن يحدث.
التفكير غير المرتبط بالإختبار	• خلال الإختبار، أفكر بالأحداث الماضية • يخترق تفكيري عدد من المعلومات الغريبة أثناء الإختبار.

	• أثناء الإختبار، أجد تفكيري مشغولا بأشياء لا علاقة لها بمادة الإختبار.
ردود الفعل الجسمية	• يصيبني صداع رأسي خلال الإختبارات الهامة • تبدو معدتي مضطربة قبل الإختبار. • تزداد ضربات القلب لدي حالما يبدأ الإختبار.

تصنيف الطلبة حسب درجة قلق الإختبار لديهم

هناك ثلاثة أنواع من الطلبة حسب درجة قلق الإختبار لديهم، وتكمن مهمة المعلم في معرفة طلابه الذين يصنفون في أحد تلك الأنواع، يساعد كثيرا في العمل معهم ليقدموا أفضل ما لديهم في التقييم:

النوع الأول هم الطلبة الذين ليس لديهم مهارات دراسية جيدة ولا يدركون الأفكار الرئيسية للموضوع الذي يدرسونه ومدى القدرة على تنظيمها، ويبدو هؤلاء الطلبة قلقين على التقييم القادم بسبب أنهم لم يتعلموا جيدا.

النوع الثاني: هم المتمكنون من المادة الدراسية جيدا ولديهم مهارات دراسية جيدة ولكنهم يخشون الفشل في التقييم.

النوع الثالث هم كل من يعتقد أن لديه مهارات دراسية جيدة ولكنه في الحقيقة غير ذلك. هؤلاء يكون أداؤهم على التقييم ضعيفا ويكونوا أكثر قلقا عند التقييم.

كيف يمكن مساعدة الطلبة الذين لديهم قلق اختبار؟

تتعلق العوامل التالية بقلق الإختبار، ويتوقع أن تقع ضمن إمكانية سيطرة المعلم وتحكمه في أوضاع التقييم الصفية:

- عندما يعتقد الطالب بأن التقييم صعب فإن قلق الإختبار لديه يرتفع.

- يعاني الطلبة ذي الأوضاع الصعبة مستوى أعلى من قلق الإختبار من الطلبة العاديين.
- يعاني الطلبة الذين يحصلون على تغذية راجعة على الإختبار فقرة بعد أخرى درجة أقل من الطلبة الذين لا يتمتعون بمثل تلك التغذية الراجعة.
- يسبب الإختبار الذي يتم ترتيب فقراته من السهل حتى الصعب قلقا أقل منه عندما يتم الترتيب بشكل آخر.
- تكرار فحص الطلبة ذوي القلق المرتفع يحسن من مستوى أدائهم.
- يعاني الطلبة ذو الأكثر قلقا من التشتت في الانتباه بالوسائل السمعية والمرئية أكثر من قليلي القلق.
- إعطاء الطلبة الذين لديهم قلق اختبار عالي تعليمات خاصة لتركيز انتباههم وعدم تشتيتهم وذلك بتكرار عبارات "لا تقلق كثيرا" و "أرجو أن تكون أحسن".
- يمكن للطلاب منخفضي قلق الإختبار أن يخففوا قلقهم بعد تلقيهم تدريبا على حكمة الإختبار.

ويمكن أن يتم توجيه أسئلة للطلبة حول مساعدتهم علي الاسترخاء ويصبحوا أقل توترا خلال وبعد التقييم. وهنا يمكن تقديم الاقتراحات التالية:

- لا يجوز للمعلم أن يتكلم أو يقاطع أثناء انشغال الطلبة في الأداء على التقييم.
- من المفروض أن يقوم المعلم بمراجعة المادة التعليمية التي سيتم تقييم الطلبة فيها قبيل موعد التقييم.

- لا ينصح بأن يكرر المعلم التجوال حول الطلبة ومن فوق أكتافهم أثناء التقييم. وكذلك فتح أحاديث جانبية معهم.
- المفروض أن يقوم المعلم بزرع جو من الثقة بأداء الطلبة على التقييم ويتجنب التلميح بأن الامتحان القادم سيكون صعبا جدا.

التصحيح لأثر التخمين Correction for Guessing

في حالة الفقرات من نوع الصواب والأخطاء أو الاختيار من متعدد، فإن معادلة التصحيح لأثر التخمين قد تستخدم أحيانا في تقدير درجات الطلبة. وهنا يتم التعامل مع تلك المعادلة بمسارات ثلاثة هي:

الأول: عندما تنص تعليمات الإجابة على الإختبار على ضرورة أن يجيب الطالب على فقرات الإختبار جميعها، على أن لا يتم ترك أي فقرة دون إجابة. وعليه، يتم تطبيق المعادلة التالية:

$$X = R - \frac{W}{n-1}$$

حيث يرمز الحرف X إلى درجة المفحوص المصححة. ويرمز الحرف R إلى عدد الفقرات التي تم الإجابة عنها إجابة صحيحة. ويرمز الحرف W إلى عدد الفقرات التي تم الإجابة عنها إجابة خاطئة. والحرف n إلى عدد البدائل التي تتكون منها الفقرة.

ففي حالة الفقرات ذات البديلين أو من نوع الصواب والخطأ، فإن n تساوي (2)، وعليه، فإن الدرجة المصححة ستكون كما يلي:

$$X = R - W$$

أما في حالة الفقرات ذات الأربع بدائل مثلا، فإن الدرجة المصححة ستكون كما يلي:

$$X = R - \frac{W}{3}$$

وجدير بالذكر أنه يتم استخدام معادلة التصحيح لأثر التخمين لتقليل استفادة المفحوص من الحصول على درجات بفضل لجوئه إلــى التخمــين بشكل صحيح.

الثاني: عندما تنص تعليمات الإجابة على الإختبار على ضرورة أن يجيب الطالب على فقرات الإختبار التي يكون متأكدا منها فقط. يتم الميل نحو مكافأة من حذف الفقرات دون أن يخاطر في الإجابة عليها أو يخمنها بأي طريقة، وذلك بإضافة جزء إلى درجة المفحوص وذلك علــى النحــو التالي:

$$X = R + \frac{O}{n}$$

حيث تعرف الدرجة هنا بالدرجة المعدلة X وليست المصححة كما ســبق. ويرمز الحرف O إلى عدد الفقرات التي تم حــذفها Omitted مــن قبــل المفحوص

الثالث: ويشمل المسارين الأول والثاني. وفيه يتم معاقبة المخمن ومكافأة الطالب الحذر الذي حذف الفقرات دون أن يخاطر في الإجابــة عليهــا أو يخمنها بأي طريقة وذلك بإضافة جزء إلى درجة المفحوص حسب المعادلة التالية:

$$X_{adjusted} = R - \frac{W}{n-1} + \frac{O}{n}$$

حيث تعرف الدرجة هنا بالدرجة المعدلة $X_{adjusted}$ وليست المصححة كمــا سبق.

تمرين: أعطي طالب امتحان يتألف من (50) فقرة من نوع الاختيار مــن متعدد بواقع أربع بدائل لكل فقرة، وقد أجاب الطالب على (38) منها إجابة صحيحة. وترك الباقي دون إجابة.

أولا: إذا نصت التعليمات على أن يجاب على كافة الفقرات دون تـرك أي منها، تستخلص درجة الطالب المصححة لأثر التخمين على النحو التـالي (على اعتبار أن الفقرات المتروكة خاطئة):

$$X = R - \frac{W}{n-1} = 38 - \frac{12}{4-1} = 38 - 4 = 34$$

ثانيا: إذا نصت التعليمات على أن يجاب على فقرات الإختبار في حالة أن يكون الطالب متأكدا تماما. عندها من المفروض مكافأة كل طالـب ملتـزم بالتعليمات ويتمتع بحذر كاف حسب المعادلة التالية:

$$X = R + \frac{O}{n} = 38 + \frac{12}{4} = 41$$

تمرين: أعطي طالب امتحان يتألف من (50) فقرة من نوع الاختيار مـن متعدد بواقع أربع بدائل لكل فقرة، وقد أجاب الطالب على (38) منها إجابة صحيحة. وكان قد أجاب على (6) منها خطأ وكان قد ترك (6) منها دون إجابة.

أولا: إذا نصت التعليمات على أن يجاب على كافة الفقرات دون تـرك أي منها، تستخلص درجة الطالب المصححة لأثر التخمين على النحو التالي:

$$X = R - \frac{W}{n-1} = 38 - \frac{12}{4-1} = 38 - 4 = 34$$

حيث تم اعتبار الفقرات التي تركها الطالب خاطئة.

ثانيا: إذا نصت التعليمات على أن يجاب على فقرات الإختبار في حالة أن يكون الطالب متأكدا تماما. عندها، من المفروض مكافأة كل طالب ملتـزم بالتعليمات ويتمتع بالدقة والحذر. وتستخلص درجة ذلك الطالب المصححة (المعدلة) لأثر التخمين على النحو التالي:

$$X_{adjusted} = R - \frac{W}{n-1} + \frac{O}{n} = 38 - \frac{6}{4-1} + \frac{6}{4} = 37.5$$

الفصل الرابع

تحليل الأداء على أدوات القياس وفقراتها

Performance Analysis on Measurement Tools and Items

- مقدمة
- التحليل الكلي لمستوى الأداء
- تحليل فقرات الإختبار
  - تحليل فقرات الإختبار معياري المرجع
    - صعوبة الفقرة
    - معامل صعوبة الفقرة الإنتقائية
    - معامل صعوبة الفقرة الصياغية
    - دلالة معامل الصعوبة
    - تمييز الفقرة
    - معامل تمييز الفقرة الإنتقائية
    - معامل تمييز الفقرة الصياغية

4

- حساسية الفقرة
- مؤشر التوافق

- تحيز الفقرة
- تحليل الفقرات وحجم العينة
- العلاقة بين معامل صعوبة الفقرة ومعامل تمييزها
- نقاط للتأمل
- أسئلة وتمارين

المخرجات التعليمية

بعد دراسة هذا الفصل يتوقع من الطالب أن يكون قادرا على أن

- يقارن بين التحليل الكلي للأداء على الإختبار وتحليل فقرات الإختبار.
- يحلل فقرات من نوع معيارية المرجع.
- يحلل فقرات من نوع محكية المرجع.
- يقارن بين آلية تحليل فقرة انتقائية وأخرى صياغية.
- يقيم نوعية فقرات الإختبار في ضوء قيم معاملي الصعوبة والتمييز.
- يصف العلاقة بين معاملي صعوبة وتمييز الفقرة.
- يتحرى مدى تحيز الفقرة أو المقياس ككل.

مقدمة

يلجأ واضعو المقاييس والإختبارات أحيانا إلى كتابة تقرير حول نتائج أداء الممتحنين على أداة القياس موضع البحث. فقد يستخدم ذلك التقرير في إعطاء صورة عن مستوى أداء المجموعة، وموقع أداء الممتحن بالنسبة لتلك المجموعة من ناحية، أو ربما يستخدم في تطوير تلك الأداة عن طريق حذف أو تعديل أو تطوير بعض الفقرات، أو ربما الإبقاء على بعضها الآخر من ناحية أخرى. كما يمكن أن يستخدم التقرير في وصف صعوبة الإختبار ككل، وكذلك تحديد مواطن ضعف وأماكن قوة أداء الممتحن، هذا خصوصا عندما يتم الحديث عن اختبارات التحصيل أو القدرات أو الاستعدادات. والجدير بالذكر أن عملية التحليل يمكن أن تصنف في مستويين هما:

الأول: التحليل على مستوى الأداة ويعرف بالتحليل الكلي، وفيه تعد الدرجة الكلية للمفحوص على الإختبار أو المقياس بمثابة المتغير الذي يخضع لعملية التحليل والتفسير. وجدير بالذكر فإنه قد يتم إيجاد علامة كلية على الإختبار أو على كل بعد من أبعاده، كما أنه من غير الممكن دائما أن يصار إلى إيجاد علامة كلية على أداة القياس بإيجاد حاصل جمع العلامات على أبعاد تلك الأداة. خاصة عندما تتفاوت أهمية كل بعد مقارنة مع أهمية الأبعاد الأخرى، أو عندما يختلف عدد فقرات كل منها.

عند ذلك، من المفروض أن يصار إلى تحويل العلامات الخام لكل بعد إلى علامات معيارية (كما في اختبار بيتا III للذكاء غير اللفظي، واختبارات ويكسلر) ثم نجد مجموع تلك الدرجات المحولة، وهكذا. من ناحية أخرى، يمكن إيجاد علامة كلية على الإختبار من مجموع الدرجات على أبعاد الإختبار إذا كانت تلك الأبعاد تتمتع بذات الأهمية في قياس السمة التي بني الإختبار من أجل قياسها. ويقتصر التحليل في هذا النوع

على عرض وتحليل المؤشرات الإحصائية الرئيسة المتعلقة بالدرجات الكلية من مثل:

- **التوزيع التكراري للدرجات وتمثيلها البياني.**
- **مدى الدرجات من خلال أكبر وأقل درجة.**
- **مؤشرات النزعة المركزية.**
- **مؤشرات التشتت.**
- **شكل التوزيع بدلالة مؤشرات النزعة المركزية والتشتت**
- **مؤشرات التفلطح والالتواء.**
- **توزيع الدرجات حسب اختلاف خصائص الممتحنين.**

الثاني: التحليل على مستوى الفقرة أو العبارة، ويدعى بتحليل الفقرات Item Analysis، وتكون فيه الدرجة على العبارة أو الفقرة هي التي تخضع للتحليل والبحث. ويتضمن هذا التحليل دراسة المؤشرات التالية:

- **صعوبة الفقرة أو شدة العبارة.**
- **تمييز الفقرة.**
- **جاذبية المموه.**
- **حساسية الفقرة للتدريس.**
- **متوسط الفقرة وتباينها.**
- **صدق الفقرة وثباتها.**
- **تحيز الفقرة.**

التحليل الكلي لمستوى الأداء

يتضمن التحليل الكلي **The Wholistic Analysis** لمستوى أداء مجموعة من الممتحنين على مقياس أو اختبار ما التعرف على مؤشرات كل من النزعة المركزية والتشتت وعلاقة كل منها بشكل توزيع درجات

الممتحنين وعلاقة ذلك بمستويات قياس المتغيرات. ويتضمن تقرير التحليل الكلي وصفا لأكبر قيمة وأقل قيمة وأكثر القيم تكرارا، كما يصف كيفية تجمع البيانات من جهة وتباعدها عن مراكز التجمع من جهة أخرى. وبناءا على هذه المؤشرات يمكن الحكم بشكل كلي على صعوبة أو سهولة أداة القياس بالنسبة للمفحوصين مشيرا إلى خصائص هؤلاء من حيث مستواهم الدراسي وفئاتهم العمرية، وغيرها من المتغيرات ذات العلاقة.

ولضمان أن يقوم مطور الأداة بعمله على أحسن ما يرام، لا بد من التعرف على مفاهيم النزعة المركزية والتشتت وأشكال التوزيعات التكرارية للبيانات بشيء من التفصيل، وعلاقة كل منها بمستويات القياس. تعرف مؤشرات النزعة المركزية بأنها نقاط تقع على مستوى قياس المتغير، وتصف موقع أو مواقع تراكم أو تجمع الدرجات في التوزيع. وهذه النقاط لا تسهم كثيرا في إعطاء معلومات عن شكل التوزيع (Bartz, 1988; Besag and Besag, 1985; Glassnapp and Poggio, 1985). فقد تجد توزيعا تتركز معظم درجاته في الوسط وقليل منها يتناثر على أطراف التوزيع، ويعرف هذا بالتوزيع الطبيعي أو الأقرب إلى الطبيعي، عند ذلك تبدو مقاييس النزعة المركزية متقاربة وتقع في مركز التوزيع.

وفي أحيان أخرى، تجد توزيعا تتركز معظم درجاته في طرف من طرفي التوزيع، ويدعى توزيعا ملتويا، ولا تبدو مقاييس النزعة المركزية متجمعة في مركز التوزيع أو قريبا منه، بل ربما تظهر بمواقع مختلفة وبالتالي يتمتع التوزيع المذكور بخصائص مختلفة تبعا لذلك. ومهما يكن من أمر، فإن أكثر مقاييس النزعة المركزية شيوعا وأهمية في وصف الدرجات وتحليلها هي المنوال (Mode)، والوسيط (Median) والمتوسط الحسابي(Mean).

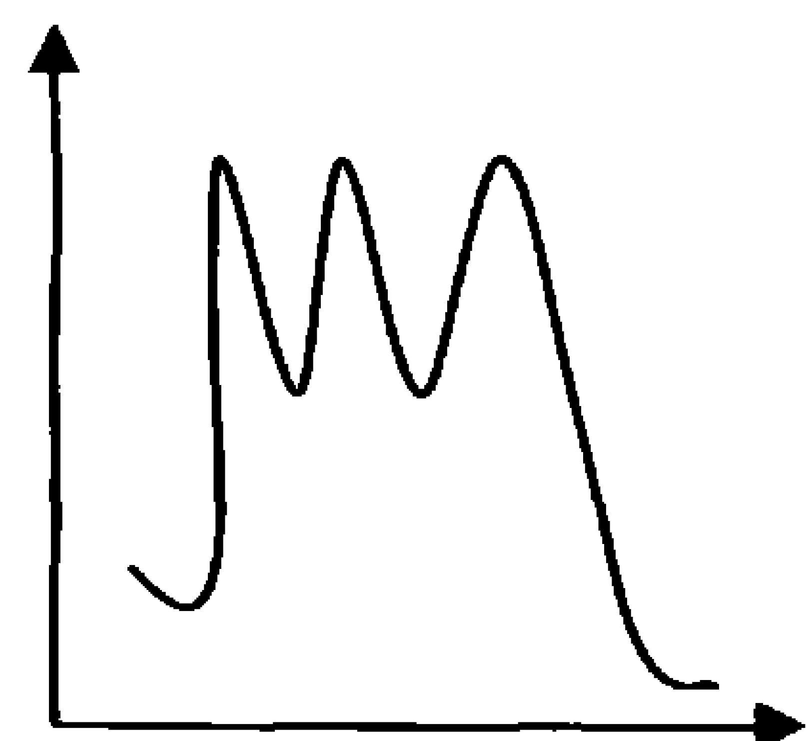
المنوال The Mode

هو أبسط مؤشر للنزعة المركزية، ويعرف بأنه القيمة الأكثر

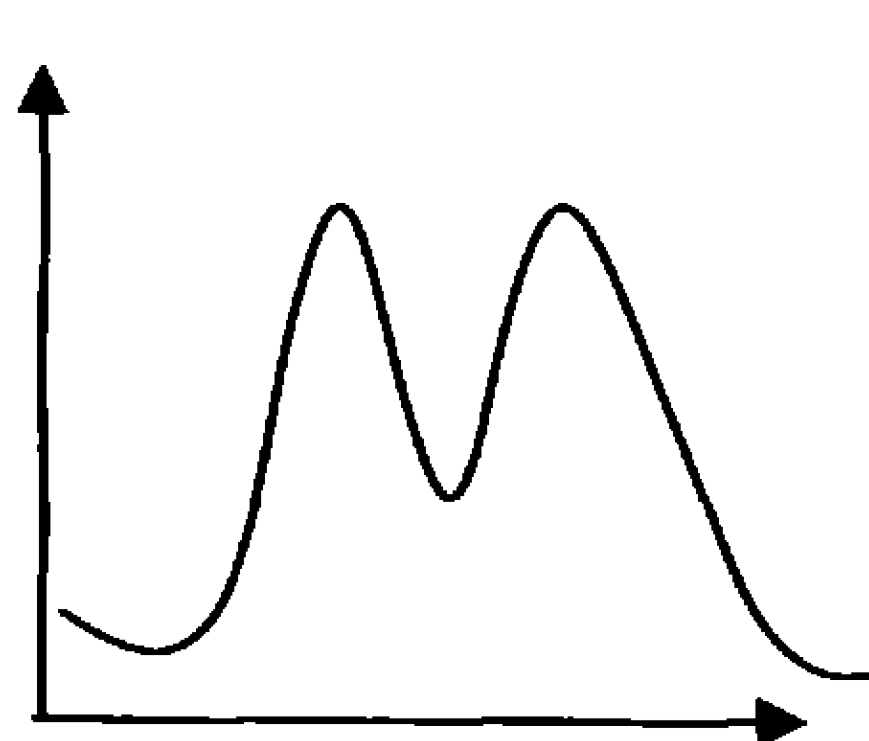
تكرارا في التوزيع. ويمكن تحديد المنوال بتفحص درجات مجموعة من الطلبة في اختبار معين، وحصر الدرجة الأكثر تكرارا في التوزيع. إذ لا يتطلب تقدير المنوال أية عمليات حسابية. فلو نظرنا إلى الدرجات (20، 40، 80، 80، 80، 90، 90، 100) تكون الدرجة (80) هي المنوال، كونها تتكرر ثلاث مرات، وهو أكبر من تكرار أية درجة أخرى في المجموعة.

ويمكن أن يستخدم **المنوال مع المتغيرات النوعية والكمية**، وهو الوحيد من بين مقاييس النزعة المركزية الذي يمكن استخدامه في وصف البيانات التي يمكن قياسها على المستوى الاسمي التصنيفي (Glass and Hopkins, 1996).

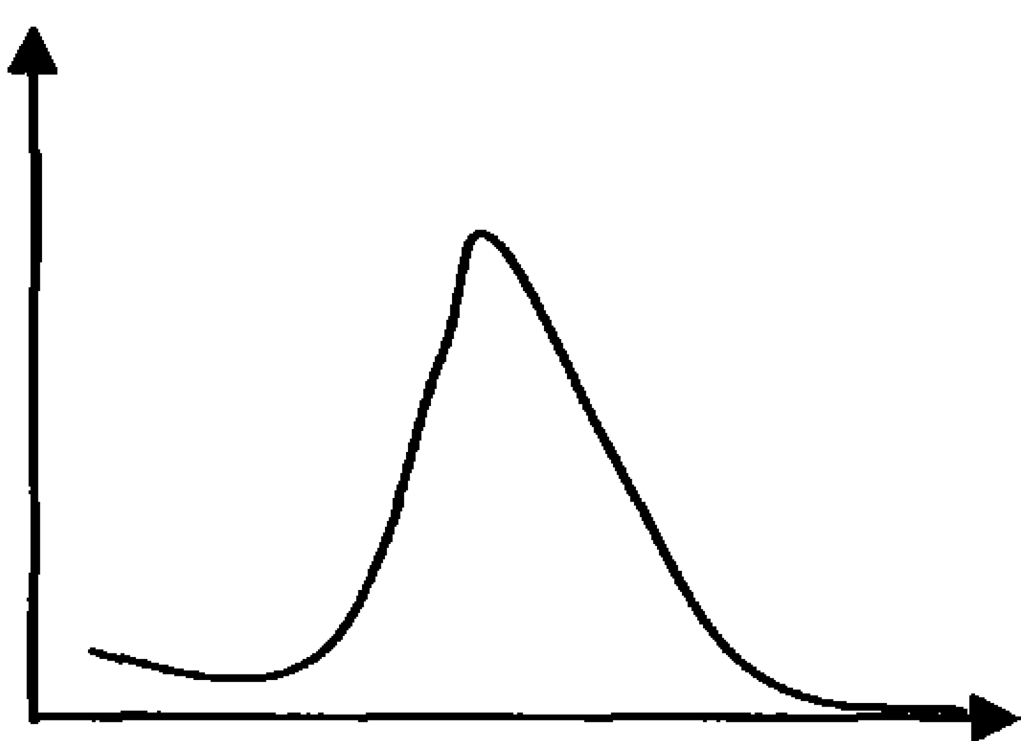
وعند التعامل مع بيانات مجدولة، يكون المنوال هو مركز الفئة ذات التكرار الأكبر. وتدعى الفئة التي تتضمن أكبر تكرارا بالفئة المنوالية Modal Interval. والجدير بالذكر بأن بعض التوزيعات المجدولة أو غير المجدولة قد تحوي منوالا واحدا كما في الشكل رقم (1/4)، وقد يحوي التوزيع منوالين ويدعى توزيع ثنائي المنوال Bi-Modal Distribution كما في الشكل رقم (2/4). وفي حالة وجود ثلاث قيم ذات تكرار أكبر، فيدعى ثلاثي المنوال أو متعدد المنوالات Multi-Modal Distribution كما في الشكل رقم (3/4):



شكل رقم (3/4)



شكل رقم (2/4)



شكل رقم (1/4)

تعد مسألة تحديد منوالين أو أكثر للتوزيع الواحد قضية جدلية. فقد يرى البعض ضرورة تساوي تكرار كل القيم المنوالية، في حين يرى البعض الآخر التساهل في هذا الأمر. ففي البيانات (**10، 11، 11، 11، 11، 11، 12، 13، 14، 14، 14، 14، 17**) تعد القيمة (**11**) منوالا ، بينما يقول البعض الآخر بأن القيمتين (11) و (14) منوالان على الرغم من أن تكرار القيمة (11) هو (5) ، بينما تكرار القيمة (14) هو (4) مرات، وفي هذه الحالة تدعى القيمة (11) بالمنوال الرئيسي Major Mode بينما القيمة (14) منوالا ثانويا Minor Mode.

وقد جاء في بعض المصادر أن القيم أو مراكز الفئات تعد منوالات إذا كانت تكراراتها هي الأكبر ومتساوية، أو إذا كان تكرار كل منها أكبر من أو يساوي 80 % من تكرار القيمة أو الفئة المنوالية الرئيسية. وفي حالة وجود منوالين أو أكثر متجاورين في حالة البيانات المجدولة أو غير المجدولة، يمكن إيجاد متوسط هذه المنوالات ليكون منوال التوزيع مساويا لقيمة ذلك المتوسط، شريطة أن تكون القيمة الناتجة موجودة في التوزيع (Hinkle, Wiersma, and Jurs, 1979). كما يمكن تفضيل المنوال على غيره من مقاييس النزعة المركزية لسهولة حسابه من ناحية، ولأنه الوحيد الذي يمكن استخدامه مع أي نوع من البيانات الاسمية والرتبية والفئوية والنسبية.

تمرين: يلاحظ أن القيم المنوالية في مجموعة البيانات الواردة في نموذج الساق والورقة التالي هي (52) و (31) :

6	6678
5	00122225
4	12346899
3	00111123

الوسيط The Median

هو تلك القيمة التي يسبقها ويليها عددان متساويان من القيم المرتبة تصاعديا أو تنازليا. وبعبارة أخرى يعرف الوسيط بأنه القيمة التي يقع تحتها أو يقل عنها 50% من القيم، وذلك بعد أن يتم ترتيبها تصاعديا أو تنازليا، وبلغة أخرى، فإن الوسيط هو المئين 50 (Cook, Grawen, and Clarke, 1999). ولإيجاد الوسيط لمجموعة من القيم، يصار بداية إلى ترتيبها تصاعديا أو تنازليا حسب قيمها، ثم نجد من بينها القيمة الوسطى عن طريق العد، فتكون هي الوسيط المطلوب كما في التمرين التالي:

تمرين: لحساب الوسيط للقيم التالية: **7، 10، 12، 9، 17، 15، 13** ، يجب ترتيب هذه القيم تصاعديا لتصبح كما يلي: **7، 9، 10، 12، 13، 15، 17** وبذلك تكون القيمة **12** هي الوسيط، لأنه يسبقها ثلاث قيم ويتبعها ثلاث قيم أيضا.

يسهل استخراج الوسيط في حالة كون عدد القيم فرديا، وذلك لوجود قيمة واحدة وسطى يمكن تحديدها بالعد. ولكن الحال يختلف إذا كان عدد القيم زوجيا، لأنه في هذه الحالة، سيكون هناك قيمتان وسيطتان وليس قيمة واحدة. وبذلك يمكن إيجاد الوسيط بحساب معدل تلك القيمتين (Glass and Hopkins, 1996) كما في التمرين التالي:

تمرين: إذا تأملنا مجموعة القيم التالية:(**5، 6، 8، 15، 17، 22، 24، 29**). تكون لها قيمتان وسيطتان هما (**15**) و (**17**) وبذلك يكون الوسيط = (16). والجدير ذكره أن الوسيط لا يتأثر بوجود قيم شاذة (متطرفة) في توزيع الدرجات وغيرها.

المتوسط The Mean

يعرف المتوسط أو الوسط الحسابي لعدد من الدرجات بأنه معدل الدرجات في التوزيع، ويحسب مقداره بقسمة مجموع القيم على عددها كما في المعادلة

التالية:

$$\bar{x} = \frac{\sum x}{n}$$

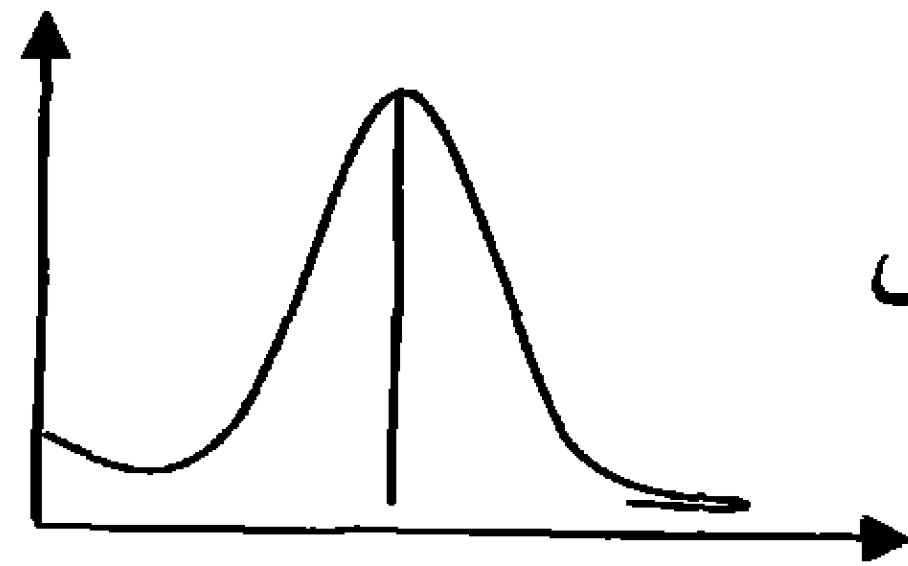
وبكل بساطة، يكون المتوسط الحسابي للقيم: 3، 4، 5، 6، 7 مساويا (5). وعند حساب مقدار الفرق بين كل من هذه القيم ومتوسطها الحسابي (−2 ، −1، صفر، 1، 2) ومن ثم إيجاد مجموعها، فيكون صفرا، وصولا إلى القاعدة المعروفة **بأن " مجموع انحرافات القيم عن متوسطها يساوي صفرا** ". الأمر، الذي يجعل المتوسط الحسابي نقطة اتزان بالنسبة لتوزيع الدرجات. وجدير بالذكر بأن الوسط الحسابي هو مؤشر النزعة المركزية المستخدم في وصف البيانات التي تقاس على المستويين الفئوي والنسبي فقط (Howell, 1999; Isaac and Michael, 1984). أما في حالة البيانات المجدولة، فيمكن حساب المتوسط الحسابي بإتباع المعادلة التالية:

$$\bar{x} = \frac{\sum fx}{\sum f}$$

حيث تمثل x القيمة أو مركز الفئة بينما ترمز f إلى التكرار. اما في حالة توزيع الساق والورقة، فيمكن حسب المتوسط بجمع القيم وقسمتها على عددها، ويمكن تجريب ذلك على البيانات الواردة في التمرين السابق. والمعروف أن المتوسط الحسابي يتأثر بكافة القيم في التوزيع. وعليه، فإن المتوسط ليس هو مؤشر النزعة المركزية الأمثل في حالة وجود قيم شاذة، بل يجب عندها استخدام الوسيط للتعبير عن هذه النزعة.

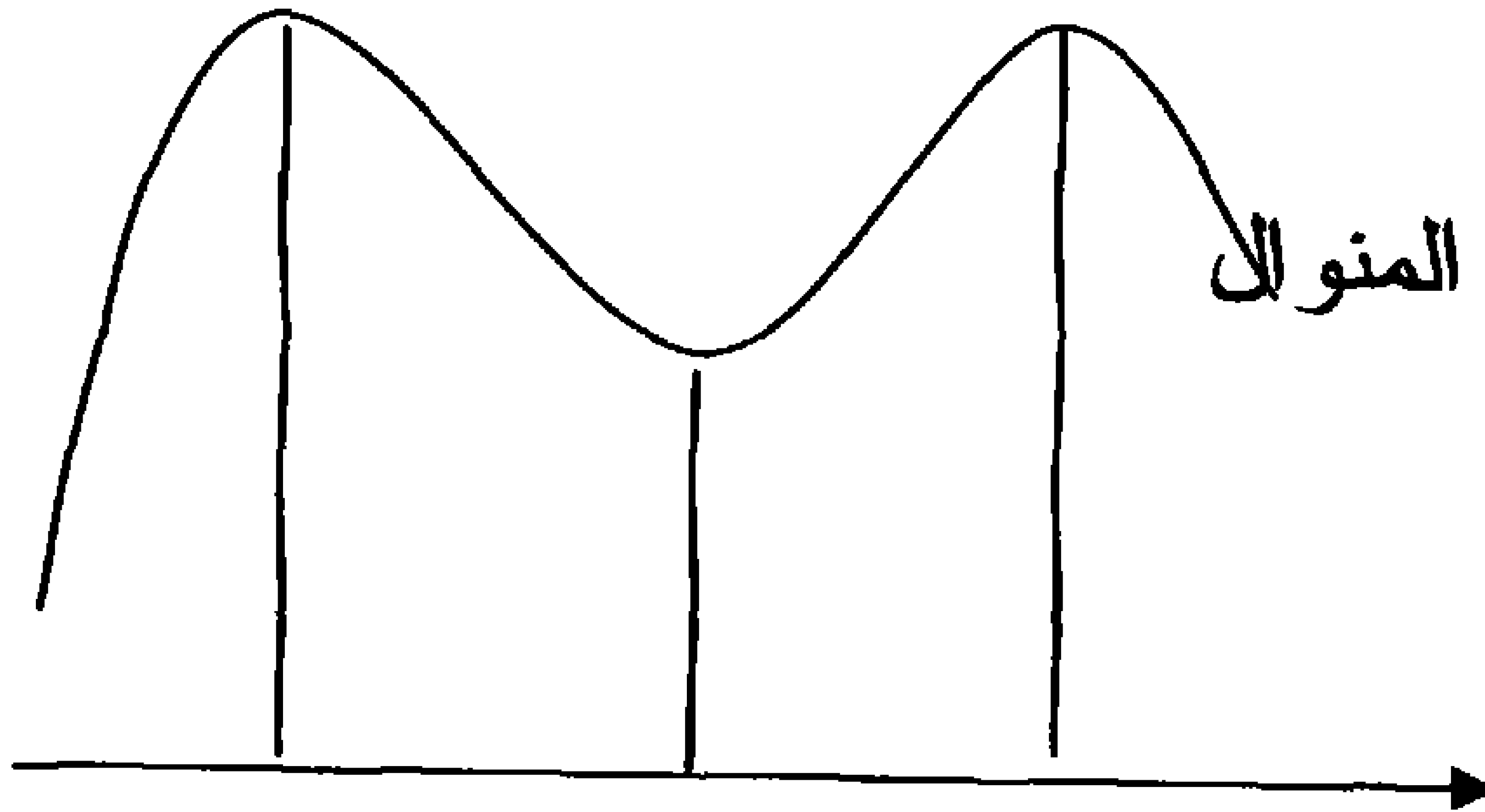
شكل توزيع الدرجات بدلالة مؤشرات النزعة المركزية

للمقارنة بين أشكال توزيع الدرجات في ضوء قيم مؤشرات النزعة المركزية يمكن الاطلاع على متضمن الأشكال من (4/4-7/4). إذ يبين الشكل رقم (4/4) شكل التوزيع المعتدل الذي يشبه الجرس ومتماثل وذو منوال واحد. والذي يعرف بالمنحنى الطبيعي الذي تتساوى فيه جميع قيم مؤشرات النزعة المركزية. وإذا كان هذا التوزيع يمثل توزيعا لدرجات مجموعة من الممتحنين على اختبار ما، فيمكن القول بأن مستوى أداء الطلبة على الاختبار معتدلا، وأن الاختبار يتمتع بصعوبة متوسطة بالنسبة لهؤلاء المفحوصين، وفيه تتساوى قيم المتوسط الحسابي والمنوال والوسيط.



الشكل (4/4): توزيع متماثل أحادي المنوال

بينما يعرض الشكل رقم (5/4) توزيعا برأسين يدعى ثنائي المنوال، ويظهر مثل هذا التوزيع عندما يطبق إختبار ذو قدرة تمييزية عالية على عدد من الطلبة ويصار إلى تصنيفهم في مجموعتين هما: مجموعة الأداء المرتفع ومجموعة الأداء المنخفض.



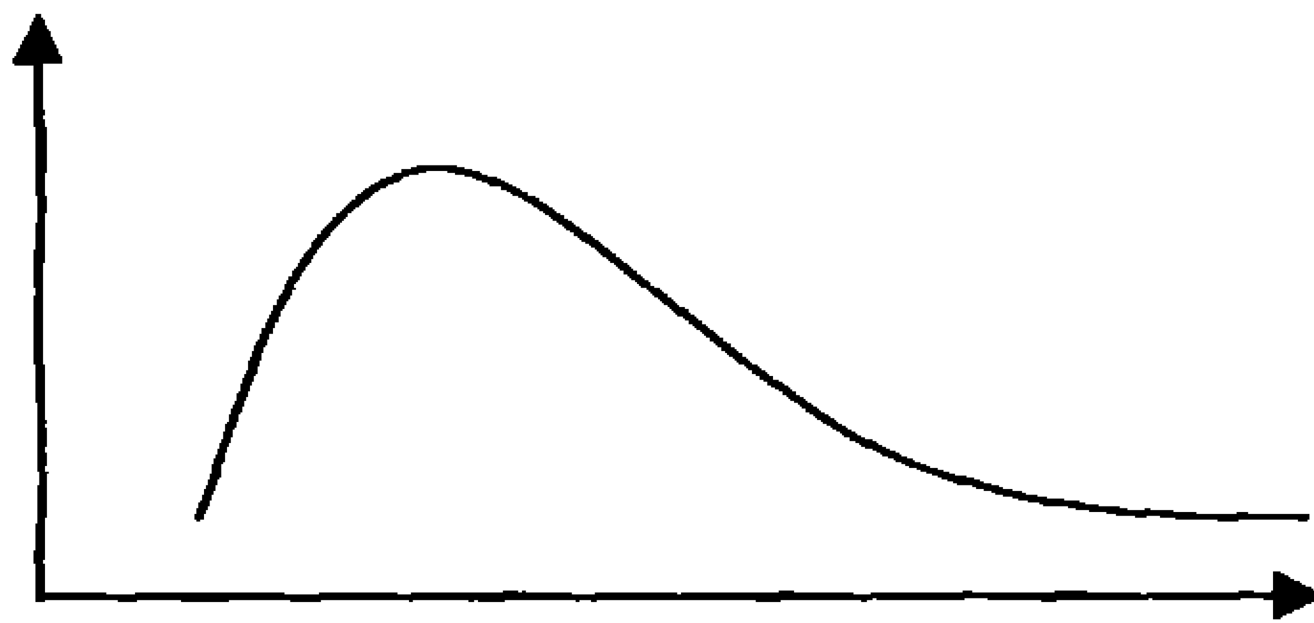
الشكل (5/4): توزيع ثنائي المنوال

كما أنه من الممكن أن يظهر التوزيع تصنيفا آخر، يوفره اختبار ذو قدرة تمييزية مختلفة بتقسيم الطلبة إلى ثلاثة مجموعات أو ربما أكثر. ويمثل الشكلان (6/4) و(7/4) توزيعين ملتويين. الأول ملتو التواءا سالبا نحو اليسار negatively skewed والثاني توزيع ملتو نحو اليمين أو ملتو التواءا موجبا positively skewed. وإذا كان كل توزيع يمثل درجات مجموعة من الممتحنين على اختبار معين، فإن التوزيع في الشكل (6/4) يمثل توزيعا لدرجات ممتحنين على اختبار سهل أو أن مستوى أداء الممتحنين على الاختبار عالي نسبيا، وتكون فيه قيمة المنوال أكبر من قيمة الوسيط هذه أكبر من المتوسط. أما الشكل رقم (7/4) فيمثل توزيعا لدرجات طلبة على اختبار صعب، أو أن مستوى أداء الممتحنين على الاختبار ضعيف نسبيا، وفيه تكون قيمة المتوسط أكبر من الوسيط وهذا أكبر من المنوال.

الشكل (6/4): توزيع ملتو نحو اليسار



الشكل (7/4): توزيع ملتو نحو اليمين



ولمزيد من إلقاء ضوء أكثر على وصف توزيع الدرجات وتحليلها بالأسلوب الكلي لا بد من التعرض إلى **مؤشرات التشتت** التي تهدف إلى كشف مدى تباعد القيم عن مراكز التجمع (المتوسط الحسابي مثلا). وهذه المؤشرات تدعى بمؤشرات مسافة، فهي موجبة دائما وليس لها وجود فعلي في التوزيع كما هو الحال بالنسبة لمؤشرات النزعة المركزية. وأهم هذه المؤشرات المدى، ونصف المدى الربعي، والانحراف المعياري، والتباين.

The Range المدى

يعتبر المدى أسهل طريقة لقياس درجة التشتت للبيانات الكمية، ويساوي مقدار الفرق بين أكبر قيمة فعلية وأقل قيمة فعلية في التوزيع (Glass and Hopkins, 1996). والمعروف أن أكبر قيمة فعلية هي أكبر قيمة ظاهرية في التوزيع مضافا إليها نصف وحدة القياس والتي تساوي (0.5) في حالة الدرجات، وأن أصغر قيمة فعلية هي أصغر قيمة في التوزيع مطروحا منها (0.5). فالمدى لدرجات خمس طالبات (40 ، 66 ، 70، 98، 81) يساوي (59) وهو حاصل طرح أعلى قيمة فعلية لتلك الدرجات (98.5) وأقل قيمة فعلية لها (39.5).

يفيد المدى في إعطاء فكرة سريعة عن تشتت البيانات في الوقت الذي لا تكون الدقة فيه ضرورة أساسية. إذ على الرغم من سهولة حساب المدى، فهو غير دقيق، لكونه يركز على قيمتين فقط (أعلى قيمة وأقل قيمة). ويتأثر تبعا لذلك بالقيم المتطرفة أو الشاذة Glassnapp and Poggio, 1985). وتتضح الصورة أكثر إذا استعرضنا مجموعات القيم التالية التي تختلف في عدد عناصرها (4)، (9)، (2) وفي قيم متوسطاتها (9)، (2.33)، (105) على الرغم من أن لها نفس مقدار المدى (13):

المجموعة الأولى: (3، 7، 11، 15)

المجموعة الثانية: (1، 1، 1، 1، 1، 1، 1، 1، 13)

المجموعة الثالثة: (99، 111).

وبالرغم من أن المدى أسهل مقاييس التشتت حسابا وأقلها دلالة لأنه يتعامل مع قيمتين فقط في التوزيع، إلا أنه يمكن أن يؤدي فائدة معقولة إذا تم استخدامه جنبا إلى جنب مع أحد مقاييس النزعة المركزية (المتوسط مثلا) للمقارنة بين توزيعين أو أكثر من البيانات، أو عند الحكم على كفاءة شخص لشغل وظيفة أو مهنة ما (زيتون، 1996).

يمكن القول أيضا بأن كفاءة المدى تقل كثيرا مع تغير طبيعة وحجم العينة. وهو مؤشـــــر مضلل (misleading) عندما يستخدم لمقارنـة مجموعتين تختلفان في الحجم، إذ أن زيادة حجم المجموعة ربما يزيد مـن احتمالية وجود قيم متطرفة، إضافة إلى أن المدى ربما لا يفيد في إعطـاء صورة عن شكل انتشار البيانات، خاصة عندما يتساوى مدى مجمـوعتين من البيانات بالرغم من اختلاف متوسط كل منهما. فربما يكـون توزيـع أحدهما ملتويا بشكل معين ويكون الآخر غير ذلك (**Glass and Hopkins, 1996; Gravetter and Wallnau, 1991**).

نصف المدى الربعي Semi- Interquartile Range

هو أحد مؤشرات التشتت الذي يساوي نصف المسافة بين الربع الثالث والربع الأول في التوزيع ، أو هو نصف الفـرق بـين الربـع الأعلى أو الثالث (Q_3) والربع الأدنى أو الأول (Q_1). وبلغة المئينات، فإن نصف المدى الربعي يساوي نـصف المـسافة بـين المئـين 75 والمئين 25. كما يمكن النظر إلى هذه المسافة على أنها متوسط بعـد الوسيط عن الربع الثالث وبعد الوسيط عن الربع الأول.

جاء استخدام نصف المدى الربعي عوضا عن المـدى، كـون الأخير يتأثر بالقيم المتطرفة والشاذة. والمعروف أن بين نصف المدى الربعي والوسيط صفات مشتركة، فكلاهما يستخدم المئينات ولا يتـأثر بالقيم المتطرفة. ولاشك أن قيمة نصف المدى الربعي تقترب كثيرا من قيمة الإنحراف المعياري. ويتم حسابه حسب المعادلة التالية:

$$Q = \frac{Q_3 - Q_1}{2}$$

تمرين: أحسب نصف المدى الربعي للقيم: 1، 4، 5 ، 7 ، 9، 10، 11، 12 وقارنه مع المدى. وكم تصبح قيمة كل من المدى ونصف المدى الربعي إذا تم استبدال القيمة (12) بالقيمة (99) ؟

الحل:

Q_1:

$(25 \div 100) \times 8 = 2$

$Q1 = 4.5$

Q_3:

$(75 \div 100) \times 8 = 6$

$Q_3 = 10.5$

$$Q = (10.5 - 4.5) \div 2 = 3$$

$$R = 13.5 - 2.5 = 11$$

وفي حالة استبدال القيمة 12 بالقيمة 99، فإن المدى يصبح 99 ، بينما يبقى نصف المدى الربعي بدون أي تغيير(3). وهذا يشير إلى أن نصف المدى الربعي لا يتغير بوجود قيم متطرفة في التوزيع كما هو الحال بالنسبة للمدى.

ويتضح مما سبق أن نصف المدى الربعي Q مقياس ترتيبي للتشتت لا يأخذ بعين الاعتبار كافة القيم، بل يأخذ القيم المحيطة بترتيب الربعين الأول والثالث. وبذلك فهو يهمل حوالي نصف القيم ولكنه بديل للمدى وخاصة في حالة وجود قيم متطرفة في التوزيع أو توزيعات ذات فئات مفتوحة (Cook, Grawen, and Clarke, 1999).

الإنحراف المعياري The Standard Deviation

بالرغم من الصعوبة النسبية في عملية حساب الإنحراف المعياري مقارنة مع مقاييس التشتت الأخرى، إلا أنه يعد من أكثر تلك المقاييس شيوعا وأهمية، لاعتماده في العديد من العمليات الإحصائية المتعلقة بإجراء المقارنات واختبار الفرضيات وتقدير الدرجات المعيارية. هذا ويعرف

الإنحراف المعياري لمجموعة من القيم بأنه الجـذر التربيعـي لمتوسـط مربعات انحرافات القيم عن متوسطها الحـسابي (Hinkle, Wiersma, and Jurs, 1979). وجدير بالذكر أن الإنحراف المعياري قيمة موجبة لا يمكـن أن يكون سالبا كونه مقياس مسافة أو بعد.

كذلك، ليس للانحراف المعياري وجود فعلي على المحور الأفقـي للتمثيل التكراري للبيانات كما هو الحال بالنسبة لمقاييس النزعة المركزيـة التي هي تمثل مقاييس نقطية، لها وجود فعلي فـي التوزيـع. ولحـساب الإنحراف المعياري يجب التمييز بين البيانات، من حيـث كونهـا تمثـل مجتمعا أو عينة.

ولحساب الإنحراف المعياري للمجتمع أو للعينة، هناك معادلة تعرف بالمعادلة العامة، حيث ينصح باستخدامها عندما يكون المتوسط الحسابي كسرا، والمعادلة التالية تستخدم في حساب الإنحراف المعياري للعينة:

$$s = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}}$$

بينما تستخدم المعادلة التالية في حساب الإنحراف المعياري للمجتمع:

$$\sigma = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{N}}{N}}$$

وتتطلب هذه الطريقة حساب مجموع اا

ويمكن للطالب إيجاد الإنحراف المعياري للقيم التالية (1، 2، 3، 4، 5، 6، 7) على اعتبارها عينة مرة، ومجتمعا مرة أخرى ونحصل على الجوابين (2.158) و(2.00) على الترتيب.

يعد الإنحراف المعياري لمجموعة من الدرجات مؤشر كمي يعبر عن مقدار تشتت هذه الدرجات عن متوسطها الحسابي ، فإذا كـان الإنحـراف

المعياري لعينة من البيانات يساوي (2)، فهذا يعني بأن تلك البيانات تختلف أو تنحرف عن متوسطها بمقدار وحدتين. ولهاتين الوحدتين نفس وحدة البيانات (زيتون، 1996؛ عدس، 1999). فمثلا، لو تأملنا القيم (2 ،3، 4، 5 ،6، 7، 8) فسيكون الإنحراف المعياري لها يساوي (2) والمتوسط = (5)، عندها يمكن القول بأن معظم البيانات تنحصر بين انحراف معياري فوق المتوسط وانحراف معياري تحت المتوسط. أي بمعنى أن القيم تنحصر بين (5+2=7) و(5−2=3). وعندما يكون الإنحراف المعياري قيمة صغيرة فهذا يدل على أن التوزيع متقارب وتتجمع بياناته قرب متوسطها.

وتبرز أهمية الإنحراف المعياري من خلال دخوله في عمليات المقارنة في مستوى أداء مجموعتين أو أكثر من الممتحنين على أداة قياس معينة من ناحية، أو مقارنة مستوى أداء مجموعة معينة من المفحوصين على أكثر من أداة من ناحية أخرى. وبشكل آخر، يدخل الإنحراف المعياري في تكوين جداول معايير الإختبارات والمقاييس النفسية والتربوية، وهذا يسهل التعرف على موقع المفحوص بين أفراد المجموعة التي طبق الإختبار عليهم.

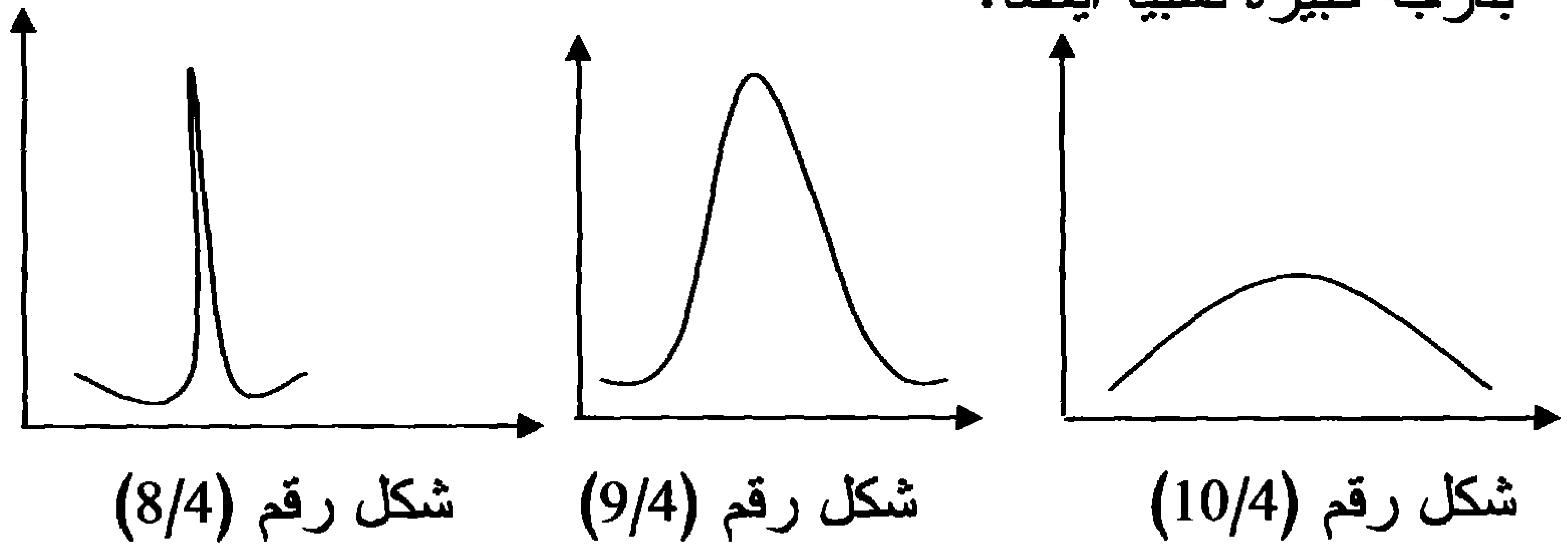
التباين The Variance

يعرف التباين بأنه معدل مربعات انحرافات القيم في التوزيع عن الوسط الحسابي، أو هو مربع الإنحراف المعياري (Glass and Hopkins, 1996). يمكن حساب التباين بإتباع الطرق التي يتم استخدامها في حساب الإنحراف المعياري بحيث يتم إيجاد مربع الإنحراف للوصول إلى قيمة التباين. ويتأثر حساب التباين كالإنحراف المعياري بكافة القيم في التوزيع بما فيها القيم المتطرفة والشاذة. الأمر الذي لا ينصح باستخدام التباين وحتى الإنحراف المعياري في حالة وجود القيم المتطرفة، بل يفضل استخدام مقياس تشتت آخر كنصف المدى الربعي (Marascuilo and Serlin, 1988; Hieman, 1992).

وعموما، فإن التباين هو الأكثر استخداما من بين مقاييس التـشتت في عمليات الإحصاء التحليلي. ولدى حساب مقدار تباين الـدرجات التـي وردت في حساب الإنحراف المعياري، وجد أن تباين تلك الدرجات = (4) في حالة اعتبارها عينة، وتساوي (4.66) في حالة كونها مجتمعا.

وعموما، فإن التباين هو الأكثر استخداما مـن بـين مقـاييس التشتت في عمليات الإحصاء التحليلي. ولدى حـساب مقـدار تبـاين الدرجات التي وردت في حساب الانحراف المعياري، وجد أن تبـاين تلك الدرجات = (4) في حالة اعتبارها عينة، وتساوي (4.66) في حالة كونها مجتمعا.

والأشكال الثلاث التالية (8/4 – 10/4) تبين التوزيـع البيـاني لنماذج تختلف في درجة تشتتها. فقد تكون درجة التشتت قليلة كما في توزيع البيانات في الشكل رقم (8/4)، وقد تكون بدرجة متوسطة نسبيا كما في الشكل رقم (9/4). و يبين الشكل (10/4) توزيعا لبيانات تتشتت بدرجة كبيرة نسبيا أيضا:



شكل رقم (10/4) شكل رقم (9/4) شكل رقم (8/4)

وبالختام، تبدو مسألة الاهتمام بمقاييس التشتت بـين درجـات المفحوصين مهمة إلى درجة كبيرة عندما تكون هذه الدرجات حصيلة أداء مجموعة من الأشخاص على أداة قياس من نوع NRT، في حين

لا تكون مسألة تشتت العلامات قضية مهمة في حالة الأداء علـــى أداة قياس من نوع CRT.

تقرير: التحليل الكلي لمستوى تحصيل طلبة على اختبار تحصيلي

تم بناء اختبار في القياس النفسي يغطي ثلاث وحـــدات دراســـية مـــن القياس النفسي في قسم علم النفس، كامتحان منتصف الفصل الأول للعـــام الجامعي 2001/2000. وبعد أن جرى تطوير مبدئي لفقرات الإختبار التي كانت في معظمها من نوع الاختيار من متعدد، تم تطبيقه على (60) فـــردا من طلبة التخصص المذكور بواقع (27) طالبا و (33) طالبة. وللتعـــرف على خصائص عامة للاختبار وكذلك الوصول إلى صورة ســـريعة عـــن مستوى هؤلاء الطلبة الممتحنين، وتحقيقا لذلك، جرى تحليل كلي لمـــستوى أداء الطلبة وذلك من خلال:

- **التوزيع التكراري للدرجات النهائية على الإختبار موضع البحث.**
- قراءة مؤشرات النزعة المركزية والتشتت لتلك الدرجات.
- التعرف على شكل توزيع الدرجات.
- تقييم دلالة اختلاف مستوى أداء الطلبة تبعا لاختلاف جنسهم.

جرى تصحيح الإختبار ومن ثم إدخال الدرجات إلى ذاكرة الحاسوب، وباستخدام برنامج **SPSS**، تم الحصول على التوزيع التكراري للـــدرجات **Score** من خـــلال الأمـــر **(Frequencies)** الـــذي يقـــع تحـــت الأمـــر **(Descriptives)** وينتمي إلى الأمر الرئيس **(Analyze)** واتضح من ذلك أن عدد الدرجات (60) وأن أقل درجة حصل عليها الطلبة هي (11) وأعلـــى درجة هي (25). كما لوحظ أن أكثر الدرجات تكـــرارا هـــي (22)، فقـــد تكررت عشرين مرة. ولم يحصل على أعلى درجة (25) سوى ممتحنـــين اثنين فقط. كما بين الجدول التكراري الثاني المتعلق بالجنس **(gender)** أن

الطلبة الذكور كانوا يشكلون 45 % من المفحوصين، بينما شكل الإناث 55 % منهم. ويمكن التعرف على معلومات أكثر من خـــلال الاطــلاع علـــى الجدولين التاليين:

Frequency Table

SCORE

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	11.00	1	1.7	1.7	1.7
	13.00	1	1.7	1.7	3.3
	14.00	1	1.7	1.7	5.0
	15.00	1	1.7	1.7	6.7
	16.00	1	1.7	1.7	8.3
	17.00	2	3.3	3.3	11.7
	18.00	3	5.0	5.0	16.7
	19.00	3	5.0	5.0	21.7
	20.00	4	6.7	6.7	28.3
	21.00	6	10.0	10.0	38.3
	22.00	20	33.3	33.3	71.7
	23.00	11	18.3	18.3	90.0
	24.00	4	6.7	6.7	96.7
	25.00	2	3.3	3.3	100.0
	Total	60	100.0	100.0	

GENDER

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	27	45.0	45.0	45.0
	2.00	33	55.0	55.0	100.0
	Total	60	100.0	100.0	

وللتعرف على مؤشرات النزعة المركزية والتشتت للدرجات الكلية للطلبة على الإختبار يمكن الاطلاع على الجدول التالي الذي تم الحصول عليه جراء الأمر frequencies الذي يحتاج إلى بعض الخيارات options التي تحدد الحصول على مؤشرات معينة.

وقد اتضح من خلال هذا الجدول أن المتوسط الحسابي للدرجات يساوي (20.98) بينما يتساوى كل من الوسيط والمنوال (22.00). وهذا يشير إلى أن توزيع الدرجات أقرب إلى الملتوى نحو اليسار أو التواؤه سالبا. وهذا يتفق مع شكل التوزيع الذي يظهر لاحقا من خلال التمثيل البياني المعروف بالمدرج التكراري Histogram. وهناك دليل آخر أيضا وهو قيمة مؤشر الالتواء Skewness التي تشير قيمتها في الجدول إلى (-1.555).

وهذا يتضح كذلك من خلال استعراض توزيع الدرجات في التوزيع التكراري، حيث أن نسبة كبيرة من الدرجات العالية ظهرت في التوزيع، في حين تقل تكرارات الدرجات المنخفضة بشكل واضح.

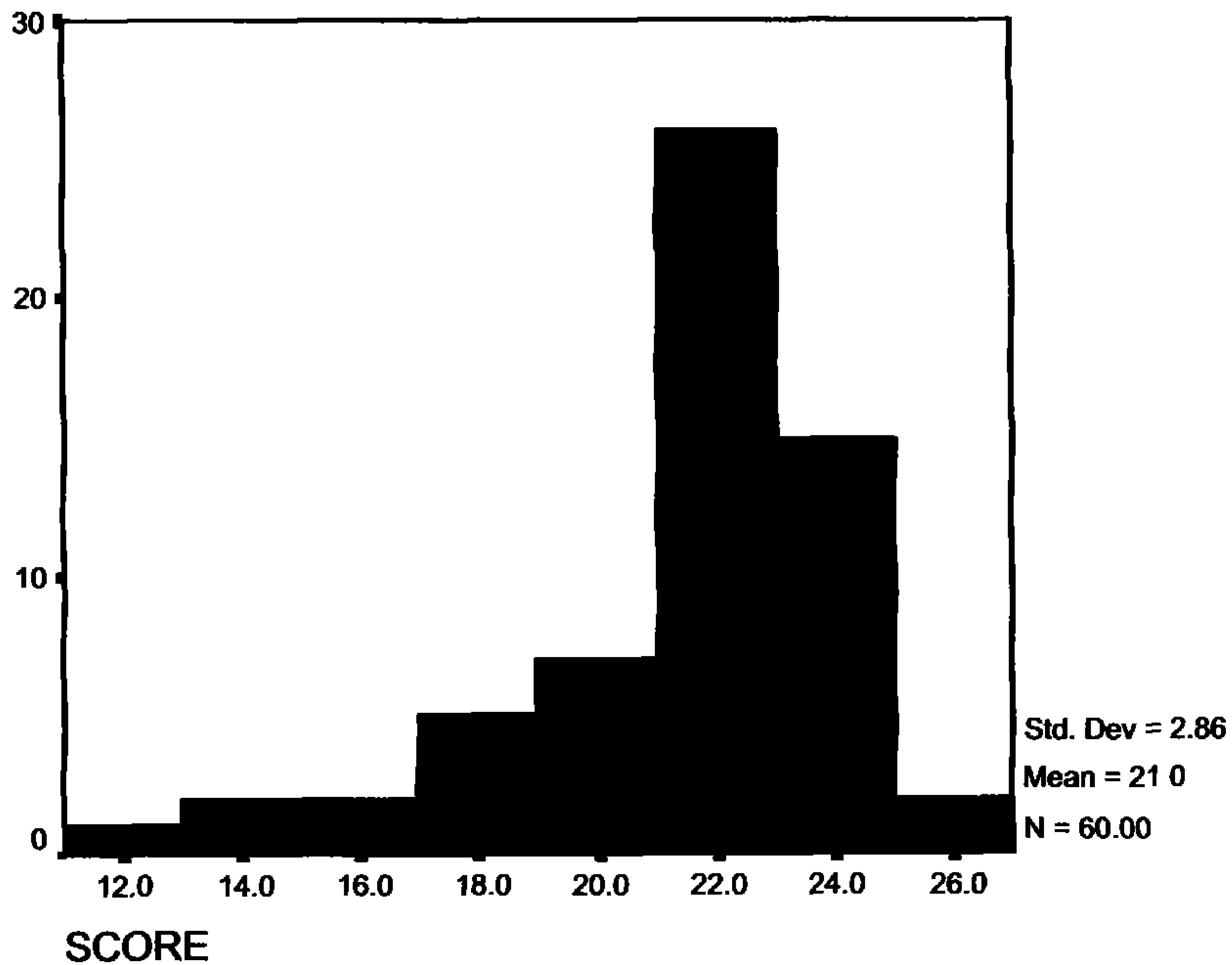
إن مجمل هذه النتائج تشير إلى أن الإختبار سهل إلى درجة معقولة وأن أداء الطلبة المفحوصين أقرب إلى العالي. وكذلك فهي نتائج متقاربة نوعا ما على الرغم من وجود بعض القيم التي تبعد عن متوسط القيم أو وسيطها. وهذا يتضح من صغر مقدار الإنحراف المعياري من ناحية وتقارب قيم المئينات والربعيات من ناحية أخرى، فقد كانت قيمة الربع الأول (20) والربع الثاني (22) والربع الثالث (23). وبهذا فإن قيمة نصف المدى الربعي تساوي (1.5).

Graph

Statistics

SCORE

N	Valid	60
	Missing	0
Mean		20.9833
Median		22.0000
Mode		22.00
Std. Deviation		2.8552
Variance		8.1523
Skewness		-1.555
Std. Error of Skewness		.309
Kurtosis		2.521
Std. Error of Kurtosis		.608
Range		14.00
Minimum		11.00
Maximum		25.00
Percentiles	25	20.0000
	50	22.0000
	75	23.0000



والجدير بالذكر أنه لا معنى لقيم المتوسط والمنوال والوسيط والإنحراف المعياري وغيرها إذا توفرت عن متغير تصنيفي (إسمي) كالجنس. إذ لا داعي لإجراء مثل هذا التحليل على ذلك النوع من المتغيرات.

وللمقارنة بين مستوى أداء الطلبة حسب جنسهم، تم حساب المتوسط الحسابي للدرجات الكلية لكل من الذكور والإناث، فكان متوسط الدرجات بالنسبة للذكور (21.37)، وبالنسبة للإناث يساوي (20.66) كما في الجدول التالي.

Means

Report

SCORE

GENDER	Mean	N	Std. Deviation
1.00	21.3704	27	2.3063
2.00	20.6667	33	3.2372
Total	20.9833	60	2.8552

ولاكتشاف ما إذا كان الفرق بين المتوسطين ذا دلالة من الأهمية، تم إجراء تحليل التباين الأحادي (ويمكن إجراء اختبار الإحصائي (t-test)، على اعتبار أن الدرجات الكلية متغيرا تابعا ومتغير الجنس متغيرا تصنيفيا (مستقلا). ومن خلال التحليل، تبين أنه لا يوجد فروق دالة إحصائيا بين متوسطي الذكور والإناث على الإختبار. وهذا واضح من خلال النظر إلى مقدار الإحصائي (F) التي تساوي (.901). بلغة أخرى، يمكن القول بأن هذا الفرق في مستويي الأداء لا ينسب لغير الصدفة.

One way

ANOVA

SCORE

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	7.354	1	7.354	.901	.347
Within Groups	473.630	58	8.166		
Total	480.983	59			

تحليل الفقرات Item Analysis

تعتبر عملية تحليل فقرات أدوات القياس على درجة عالية من الأهمية لما تؤديه من فوائد تساعد في الخروج بأدوات قياس فعالة تعمل على قياس السمات والصفات الإنسانية قياسا دقيقا. كما يتوقع أن يعمل تحليل الفقرات على تطوير فقرات الإختبار إلى الحد الذي يجعلها تسهم إسهاما ذا دلالة فيما يقيسه ذلك الإختبار. وبالتالي، فإن هذه العملية تنمي مهارات المشتغلين بأدوات القياس عموما وبالإختبارات على وجه الخصوص (Brown, 1976 ; Crocker and Algina, 1986).

ذلك لأن الهدف الرئيس لعملية بناء الإختبارات يتلخص في الخروج بأقصر اختبار ممكن ويتمتع بمؤشرات صدق وثبات حسب الهدف الذي يتوقع من الإختبار قياسه. وهذا لن يتحقق إلا بالتطبيق العملي لمجموعة كبيرة نسبيا من فقرات الإختبار، ويتبعها اختيار مجموعة من تلك الفقرات التي تعظم صفتي صدق وثبات ذلك الإختبار. إن مجمل عملية بناء اختبار جديد أو تطوير اختبار موجود تعرف بعملية تحليل الفقرات. وخلاصة الأمر، يمكن تعريف تحليل الفقرات بأنها العملية التي تتعلق باستقصاء الخصائص الإحصائية لاستجابات الممتحنين على كل فقرة من فقرات أداة القياس (الإختبار).

وتتضمن عملية تحليل الفقرات فحصا نوعيا وآخر كميا أو إحصائيا لخصائصها. إذ يتضمن الفحص النوعي للفقرة عملية الحكم على مدى انتمائها للمجال السلوكي أو المحتوى الدراسي الذي تمثله من ناحية، ومدى مطابقة الفقرة لقواعد الصياغة حسب نوعها من ناحية أخرى. أما الفحص الكمي، فيتضمن تقييم الخصائص الكمية التالية:

- صعوبة الفقرة **Item Difficulty** ويرمز لها بالرمز **P**
- شدة العبارة **Statement Frequency**
- تمييز الفقرة **Item discrimination** ويرمز لها بالرمز **D**
- جاذبية المموهات **Distracter Attractiveness** ورمزها **DA**
- تحيز الفقرة **Item Bias**
- بعض الخصائص الكمية والنوعية كمتوسط الفقرة، وتباينها، وصدقها وثباتها.

تختلف عملية تحليل الفقرات باختلاف نوع الإختبار: معياري المرجع **NRT** أو محكي المرجع **CRT**، أو حسب إتباع نظرية القياس الكلاسيكية **Classical Test Theory (CTT)** أو نظرية السمة الكامنه **Latent Trait Theory** التي تعرف أحيانا بنظرية استجابة الفقرة **Item Response Theory**، والرمز الشائع لها هو (**IRT**). وسيتم التعرض هنا إلى تحليل الفقرات بدلالة النظرية الكلاسيكية في القياس. وعليه، يمكن معالجة تحليل الفقرات أولا حسب اختبارات **NRT**، ثم حسب اختبارات **CRT** على النحو التالي:

I. تحليل فقرات الإختبار معياري المرجع

عرفنا سابقا أن مسألة المقارنة بين الطلبة أو الممتحنين عند جلوسهم لاختبار ما يعد أمرا مهما عندما يكون الإختبار من نوع NRT وبهذا تصبح

مسألة قدرة الفقرات على تصنيف الطلبة وترتيبهم غاية في الأهمية. كما أن الفقرة الجيدة التي تصلح لهذا النوع من الإختبارات هي تلك التي تسهم في إبراز الفروق وتعظيمها بين المفحوصين. أو هي تلك الفقرة التي تصنع تباينا ملحوظا بين مستويات المفحوصين، ويختلف هذا بالنسبة للفقرة إذا كانت في اختبار من نوع CRT. ويتضمن تحليل الفقرات التفصيلي فحص مؤشرات صعوبتها، وتمييزها، وجاذبية مموهاتها ودرجة تحيزها وغيرها.

صعوبة الفقرة Item Difficulty

يقصد بصعوبة الفقرة نسبة الطلبة الذين أجابوا الفقرة إجابة صحيحة. أو هي حاصل قسمة عدد الطلبة الذين أجابوا الفقرة إجابة صحية على العدد الكلي الذي خضع للاختبار (أو العدد الذي حاول الإجابة على تلك الفقرة). فإذا كان عدد الممتحنين (100) طالب مثلاً، وأجاب على الفقرة الأولى (20) طالباً إجابة صحيحة، فإن معامل صعوبة هذه الفقرة سيكون (.20)، أما إذا أجابها (90) طالباً إجابة صحيحة، فإن معامل صعوبتها سيكون (.90).

وفي هذا الحالة تكون الفقرة ذات معامل الصعوبة (.20) أكثر صعوبة من الفقرة التي معامل صعوبتها (.90). إذ أن الفقرة التي يجيبها إجابة صحيحة عدد قليل من المفحوصين تكون أصعب نسبيا من تلك التي يجيبها عدد أكبر من هؤلاء المفحوصين وهكذا.

إذن، كلما زادت قيمة معامل صعوبة الفقرة المحسوبة (P) بهذه الطريقة، كلما دل ذلك على سهولة الفقرة. فإذا كان معامل صعوبة فقرة ما (1) فهذا يعني أنها تامة السهولة وأن جميع الطلبة قد أجابوا عليها بشكل صحيح. أما إذا كان معامل صعوبة تلك الفقرة (صفراً) فإن هذا يعني أن الفقرة تامة الصعوبة، لدرجة أنه لا أحداً من المفحوصين قد أجابها بشكل صحيح. **من هنا يفهم بأن معامل صعوبة الفقرة هو نسبة، وعليه فإن قيمها تنحصر بين الصفر و الواحد.**

والفقرة التي يجيب عليها جميع المفحوصين إجابة صحيحة (P=1) أو تلك التي لا يجيب عليها أي منهم إجابة صحيحة (P=0) لا تفيد في الكشف عن الفروق بين المفحوصين في امتلاكهم للصفة أو السمة التي يتوقع من الإختبار قياسها. لذلك فقد أشارت الدراسات إلى أن معامل الصعوبة للفقرات التي يتكون منها اختبار معياري المرجع يكون حول (.50) هو الأفضل من حيث قدرته على جعل الفقرة تميز إلى أقصى حد ممكن بين المفحوصين.

وفيما يلي عرض لعدد من التمارينالمتعلقة بحساب معامل صعوبة الفقرة حسب آلية تصحيحها. إذ سيتم تناول معامل صعوبة الفقرة الانتقائية ثم معامل صعوبة الفقرة الصياغية (الإنشائية):

معامل صعوبة الفقرة الانتقائية

وهنا يتم التعامل مع هذا النوع من الفقرات بصرف النظر عن شكلها (الخطأ والصواب، الاختيار من متعدد، أو المزاوجة). إذ يتم تصحيحها بواقع (1) للإجابة الصحيحة أو (0) للإجابة الخطأ. ولا مجال القول بأن تلك الفقرة تستحق علامة خارج هذا المدى أو حتى قيما كسرية. وعليه، يمكن حساب معامل صعوبة ذلك النوع من الفقرات حسب الحالات التالية:

الحالة الأولى: عندما تتوفر بيانات إجمالية عن إجابة الفقرة

تمرين: ما معامل صعوبة فقرة عرضت على (40) طالبا، أجابها (10) فقط منهم إجابة صحيحة؟

الحل: **معامل الصعوبة يساوي حاصل قسمة عدد الذين أجابوا الفقرة إجابة صحيحة على العدد الكلي. وعليه، يكون معامل الصعوبة هو حاصل قسمة (10) على (40) كما في المعادلة التالية:**

$$p = \frac{10}{40} = .25$$

الحالة الثانية: عندما تتوفر بيانات تبين إجابات كل ممتحن على كل فقرة

تمرين: فيما يلي عرض لأداء مجموعة ممتحنين على اختبار يتألف من 10 فقرات من نوع الاختيار من متعدد، وقد ظهرت درجات الممتحنين على كل فقرة بواقع (0) أو (1). وقد تم حساب معامل الصعوبة (P) للفقرات العشر بحساب نسبة الإجابات الصحيحة بالنسبة لعدد الممتحنين (14). أي أنه تم إيجاد حاصل قسمة عدد من أجاب على الفقرة إجابة صحيحة على عدد الممتحنين الكلي. وقد ظهرت قيم تلك المعاملات في السطر الأخير من الجدول:

رقم المفحوص	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
01	1	1	1	1	1	1	1	0	1	1
02	1	1	1	1	1	1	1	0	1	1
03	1	1	1	1	1	1	1	0	1	0
04	1	0	0	1	1	1	1	0	1	1
05	1	1	1	0	1	1	0	1	0	0
06	0	0	0	1	1	1	0	1	0	1
07	0	1	1	1	0	0	0	1	1	0
08	1	1	1	1	0	0	0	0	0	0
09	0	0	0	0	1	0	0	1	1	1
10	1	0	0	1	1	0	0	0	0	0
11	1	1	0	0	0	0	0	0	0	1
12	0	1	0	0	0	0	0	1	0	0
13	0	0	0	0	0	0	1	0	0	0
14	0	1	0	0	0	0	0	0	0	0
P	**.57**	**.64**	**.43**	**.57**	**.57**	**.43**	**.36**	**.36**	**.43**	**.43**

الحالة الثالثة: عندما تتوفر إجابات أفراد المجموعتين الطرفيتين على الفقرة

عندما تتوفر بيانات عن مجموعة كبيرة من إجابات الممتحنين على عدد من الفقرات ثنائية التصحيح (0) و (1)، يتم بداية ترتيب الممتحنين تنازليا حسب المجموع الكلي لعلاماتهم على الإختبار ككل. ويتم استخراج أعلى 27 % من أفراد المجموعة وتعرف تلك المجموعة بـ (مجموعة

الأداء المرتفع) وكذلك أدنى27 % من أفراد المجموعة وتعرف بـ (مجموعة الأداء المنخفض). بعد ذلك يتم استخدام المعادلة التالية:

$$P = \frac{T_u + T_l}{2(n)(s)}$$

حيث يشير كل من T_u و T_l إلى مجموع علامات المفحوصين الذين أجابوا الفقرة إجابة صحيحة من أفراد مجموعة الأداء المرتفع ومجموع علامات أفراد مجموعة الأداء المنخفض على الترتيب. بينما تشير n إلى عدد أفراد مجموعة الأداء المرتفع أو مجموعة الأداء المنخفض، وتدل S على علامة الفقرة وهي تساوي (1) في حالة الفقرة الانتقائية من نوع الصواب والخطأ أو الاختيار من متعدد.

تمرين: فيما يلي عرض لأداء مجموعة ممتحنين (14) على اختبار يتألف من 10 فقرات من نوع الاختيار من متعدد، وقد ظهرت درجات الممتحنين على كل فقرة بواقع (0) أو (1). وقد تم حساب معاملات صعوبة الفقرات (P) كما ظهرت في السطر الأخير من الجدول:

رقم المفحوص	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	الدرجة الكلية
01	1	1	1	1	1	1	1	0	1	1	9
02	1	1	1	1	1	1	1	0	1	1	9
03	1	1	1	1	1	1	1	0	1	0	8
04	1	0	0	1	1	1	1	0	1	1	7
11	1	1	0	0	0	0	0	0	0	1	3
12	0	1	0	0	0	0	0	1	0	0	2
13	0	0	0	0	0	0	1	0	0	0	1
14	0	1	0	0	0	0	0	0	0	0	1
P	**.63**	**.75**	**.38**	**.50**	**.50**	**.50**	**.63**	**.13**	**.50**	**.50**	

إذ تم حساب معامل صعوبة الفقرة الأولى كالتالي:

$$P = \frac{T_u + T_l}{2(n)(s)} = \frac{4+1}{2(4)(1)} = 0.63$$

من مطالعة النتائج في هذا التمرين ومقارنتها مع النتائج في التمرين السابق، يمكن ملاحظة الفروق في قيم معاملات صعوبة الفقرات.

الحالة الرابعة: عندما تتوفر بيانات الإجابة على الفقرة على هيئة بطاقة

في هذه الحالة، تعطى بيانات الإجابة على الفقرة من قبل أفراد المجموعتين الطرفيتين (مجموعة الأداء المرتفع ومجموعة الأداء المنخفض). ومن ثم يتم تطبيق المعادلة:

$$P = \frac{T_u + T_l}{2(n)(s)}$$

تمرين: يبين الجدول التالي توزيعا لأعداد 50 مفحوصا على بدائل الفقرة من نوع الاختيار من متعدد بواقع 4 بدائل، وكان البديل (ب) هو الجواب الصحيح، والمطلوب حساب معامل صعوبة تلك الفقرة.

البديل	أعلى 27 % (14 مفحوصا)	أدنى 27 % (14 مفحوصا)
أ	صفر	2
ب*	10	3
ج	3	صفر
د	1	9

الحل: لحساب معامل صعوبة تلك الفقرة، يجب التعرف أساسا على الجواب الصحيح وهو هنا البديل (ب)، ومن ثم يتم التعويض في المعادلة كما يلي:

$$P = \frac{T_u + T_l}{2(n)(s)} = \frac{10+3}{2(14)(1)} = \frac{13}{28} = 0.46$$

معامل صعوبة الفقرة الصياغية (الإنشائية)

من المعروف أن معظم الفقرات الإنشائية يتم تصحيحها بمراعاة المعرفة الجزئية للمفحوصين. بمعنى أن الفقرة لا تصحح بصيغة ثنائية (صح =1 أو خطأ = صفر) بل أن الفقرة من هذا النوع تصحح بطريقة تختلف عن تصحيح الفقرات الانتقائية. فقد يتم تصحيح الفقرة من (10) نقاط، أو من (8)، أو من خمسة أو غير ذلك كما مر في الفصل السابق عندما تحدثنا عن قواعد التقدير Rubricks وهكذا. عندها، يمكن حساب معامل صعوبة الفقرة بطريقة المجموعتين الطرفيتين، وذلك بترتيب العلامات الكلية للممتحنين ترتيبا تنازليا. ثم يتم فرز المجموعتين الطرفيتين بواقع أعلى 27 % وتعرف بمجموعة الأداء المرتفع، وأدنى 27 % وتعرف بمجموعة الأداء المنخفض. ثم يتم التعرف على علامات الممتحنين على الفقرة المنوي حساب معامل صعوبتها في كل مجموعة، وتسجيلها، ومن ثم إيجاد مجموع كل منها. بعد ذلك يتم حساب معامل الصعوبة بتطبيق المعادلة التالية:

$$p = \frac{T_u + T_l}{2(n)(s)}$$

من ناحية أخرى، لا يختلف تفسير معامل الصعوبة للفقرة الإنشائية عنه للفقرة ذات الإجابة المنتقاة. فهي نسبة تتراوح قيمها بين (صفر) و (1). وتزداد صعوبة الفقرة كلما قل مقدار المعامل p وتكون الفقرة سهلة كلما زاد مقدار هذا المعامل، أو اقترب أكثر من (1).

تمرين: يمثل التوزيع التالي علامات الممتحنين من المجموعتين العليا والدنيا على فقرة إنشائية كان مخصصا لها 10 درجات. والمطلوب حساب معامل صعوبتها:

	الممتحنون	علامات الممتحنين	مجموع العلامات (T)
مجموعة الأداء المرتفع	أحمد، سعيد، حمد، محمد، سيف، راشد، عبدالله، جاسم	10، 5، 8، 7، 5، 7، 9، 5	56
مجموعة الأداء المنخفض	فيصل، سالم، علي، طارق، صقر، طالب، سلطان، مبارك	4، 2، 9، 10، 3، 1، صفر، 1	30

ويمكن الاستعاضة عن كتابة أسماء الممتحنين بأرقامهم المتسلسلة أو الأرقام الجامعية. كما يمكن استخدام أي رمز يسهل على مطور الإختبار التعامل معه.

الحل: نعوض القيم $(S = 10)$، $(n = 8)$، $(T_l = 30)$، $(T_u = 56)$ كما يلي:

$$p = \frac{T_u + T_l}{2(n)(s)} = \frac{56 + 30}{2(8)(10)} = \frac{86}{160} = 0.54$$

ويبدو أن هذه الفقرة تمتاز بمستوى صعوبة متوسطة. كما يجدر الإشارة هنا إلى أن طريقة حساب معامل صعوبة الفقرة الإنشائية لا تختلف سواء كانت الفقرة تتبع لمقياس من نوع NRT أو CRT.

معامل الصعوبة والجماعة المرجعية

المفروض أن يتم تحديد المجموعة المرجعية التي حسبت قيمة معامل صعوبة الفقرة بالاعتماد عليها. إذ أن صعوبة الفقرة تعتمد على خصائص المجموعة وليس صفة مطلقة من خصائص الفقرة. إذ قد تكون الفقرة صعبة بالنسبة لمجموعة الصف الخامس، وقد تكون غير ذلك بالنسبة لطلبة الصف السادس. إضافة إلى أنه من الواجب عرض الخصائص الأخرى

للمجموعة كالسنة أو المرحلة الدراسية أو الصف الدراسي أو الفئة العمرية وغيرها (Brown, 1972; Nitko, 1983).

معامل صعوبة الفقرة والمتوسط الحسابي

عرفنا بأن معامل صعوبة الفقرة هو مؤشر إحصائي يعبر عن خصائص الفقرة الموضوعية التي تصحح ثنائيا بواقع صحيحة تماما (وتستحق: **1**) أو غير صحيحة تماما (وتستحق صفرا)، بذلك فإن معامل الصعوبة يساوي عدد الدرجات (**1**) أو هي المتوسط الحسابي لدرجات المفحوصين على تلك الفقرة (**Nitko, 1983**).

وبعد أن يتم إدخال إجابات الممتحنين على فقرات أداة القياس إلى ذاكرة الحاسوب بعد تصحيحها (أي تصبح علامات ثنائيه: **0** ، **1**)، يمكن حساب متوسط الدرجات على كل فقرة باستخدام الأمر (**Analyze**) في برنامج SPSS، ثم يتم اختيار **descriptive statistics** ثم يتم اختيار **Descriptives** ومنها يتم اختيار **Mean** من زر الخيارات (**options**) والقيم الناتجة تكون هي معاملات الصعوبة للفقرات.

معامل الصعوبة ومستويات القياس

تقع معاملات صعوبة الفقرات على مستوى القياس الرتبي (Thorndike 1985). إذ يمكن تصنيف الفقرات التي صعوباتها (.67)، (.60)، (.53)، (.46)، (.39) على أنها **قيم رتبية**. إذ يمكن القول بأن الفقرة الأولى هي الأسهل يليها في ذلك الفقرة الثانية، ثم الثالثة، ثم الرابعة، وأصعبها الفقرة الخامسة. بينما لا يجوز القول بأن الفرق بين معاملي صعوبة الفقرتين الأوليتين (.67) و (.60) يساوي الفرق بين معاملي صعوبة الفقرتين الثالثة والرابعة (.53) و(.46). وهذا يشير إلى عدم تساوي وحدات معاملات الصعوبة. فهي لا ترقى إلى مستوى القياس الفئوي أو النسبي(**Nunnally, 1978; Crocker and Algina, 1986**).

وعند افتراض أن السمة التي يقيسها الإختبار تتوزع توزيعا اعتداليا، يمكن تحويل النسب المئوية من المفحوصين الذين يجيبون على فقرات الإختبار إجابة صحيحة إلى مستوى القياس الفئوي ذي الفترات المتساوية. عند ذلك يظهر توزيع معاملات الصعوبة مشابها للتوزيع الإعتدالي المعروف. فمثلا إذا كانت قيمة معامل صعوبة الفقرة يساوي (.50) فإن ذلك يعني أن 50 % من مجموعة المفحوصين المعيارية أجابوا الفقرة إجابة صحيحة. وهذه القيمة تناظر الدرجة المعيارية صفر على المحور الأفقي. ويمكن تطبيق هذا المفهوم على نماذج أخرى من معاملات الصعوبة وإيجاد ما يقابلها من الدرجات المعيارية على التوزيع الإعتدالي (علام، 2000).

معامل الصعوبة ودرجة الارتباط بين الفقرات

يجدر الإشارة إلى علاقة صعوبة الفقرة بدرجة الارتباط بين مفردات الإختبار، إذ كلما زاد معامل الارتباط بين الفقرات، لا بد من أن يكون مدى صعوبة الفقرات أكثر اتساعا مثل (.35 – .65). وفي حالة ضعف الارتباط بين فقرات الإختبار، فيتوقع أن يعتدل مدى معاملات صعوبة الفقرات (.40 – .60) (Brown, 1972; Mehrens and Lehman, 1991; Nunnally, 1978). والجدير بالذكر أن هناك شرطان رئيسان من الواجب اعتبارهما عند حساب معامل الصعوبة هما:

أهمية دراسة معامل الصعوبة

يعمل معامل صعوبة الفقرة على التحكم بمقدار تباينها، حيث أن تباين الفقرة يساوي حاصل ضرب معامل صعوبة الفقرة P بالمقدار q الذي يعرف أحيانا بمعامل السهولة (أي نسبة الممتحنين الذين يجيبون عن الفقرة

إجابة خاطئة، أي أن (p+ q =1)، أو أن (q =1−p) كما في المعادلة التالية:

$$\sigma_i^2 = p_i q_i$$

حيث يشر الرمز σ_i^2 إلى تباين الفقرة. فإذا افترضنا ثبات معامل الارتباط بين فقرات الاختبار، فإن تباين الفقرة يصل أقصاه عندما يكون معامل صعوبتها (.50) وهذا يجعل تباين الفقرة في هذه الحالة يساوي (.25)، هذا بالرغم من أن معاملات الصعوبة في معظم اختبارات الاستعداد والتحصيل المقننة معيارية المرجع تتراوح بين (.60− .80) (Crocker and Algina, 1986).

ويمكن تفسير ذلك بأن التباين الحقيقي True Variance لفقرات الاختبار يبلغ أقصى درجاته عندما يجيب على الفقرة 50 % من المفحوصين إجابة صحيحة اعتمادا على معرفتهم الفعلية بالجواب الصحيح. ومن بين هؤلاء الذين لا يعرفون الجواب الصحيح، هناك $(\frac{A}{.5})$ يمكنهم الحصول على الإجابة الصحيحة بالتخمين العشوائي حيث يشير الرمز A إلى عدد بدائل الفقرة كما في المعادلة التالية:

$$p = 0.50 + \frac{0.50}{A}$$

وتشير هذه المعادلة إلى أن P تقترب من المقدار (50.) كلما زاد عدد بدائل الفقرة.

إذ يظهر أن معامل صعوبة الفقرة ذات البديلين عند مراعاة أثر التخمين يساوي (75.) ، بينما يكون هذا المعامل (60.) للفقرة بخمسة بدائل، و(625.) لتلك التي تتألف من أربعة بدائل. ولتصحيح معامل صعوبة الفقرة لأثر التخمين نتبع المعادلة التالية:

$$p' = p - \frac{1-p}{A-1}$$

حيث يشير الرمز P' إلى معامل صعوبة العشوائي، بينما يشير الرمز P إلى معامل الصعوبة قبل التصحيح، ويرمز الحرف A إلى عدد البدائل.

تمرين: يصبح معامل صعوبة الفقرة التي أجابها 80 ممتحنا إجابة صحيحة من أصل 100 شخص حاول حلها إذا تم تصحيحها لأثر التخمين العشوائي وكانت من نوع الاختيار من متعدد بواقع خمسة بدائل:

$$p' = p - \frac{1-p}{A-1} = 0.80 - \frac{1-0.80}{5-1} = 0.75$$

أما لو كانت الفقرة من نوع الصواب والخطأ فيصبح معامل صعوبتها بعد التصحيح لأثر التخمين العشوائي (60.) كما يلي:

$$p' = p - \frac{1-p}{A-1} = 0.80 - \frac{1-0.80}{2-1} = 0.60$$

وهذا إنما يشير إلى أنه كلما زاد عدد البدائل قلت فرصة النجاح بالتخمين على اعتبار ثبات العوامل الأخرى. كما أن عددا غير قليل من التربويين يفضل التعامل مع صعوبة الفقرة قبل تصحيحها لأثر التخمين لتفضيلهم قيما كبيرة لمعامل الصعوبة.

دلالة معامل الصعوبة

نعرف أن معامل صعوبة الفقرة هو نسبة، وتتراوح قيمها بين الصفر والواحد الصحيح. وعليه يمكن فحص دلالة تلك النسبة باختبار مدى ابتعادها عن قيمة محكية (0.50 مثلا). وللقيام بذلك، يتم استخدام اختبار (Z) حسب المعادلة التالية:

$$Z = \frac{p - \pi}{S_P} = \frac{p - \pi}{\sqrt{\frac{pq}{n}}}$$

حيث يشير الرمزان P و π إلى معامل الصعوبة المحسوب للفقرة، ومعامل الصعوبة الافتراضي أو المنوي فحص الدلالة بالنسبة إليه على الترتيب. ويشير الرمز S_P إلى الخطأ المعياري للنسبة. أي أنه إذا كانت p تمثل القيمة المحسوبة لمعامل صعوبة الفقرة، فإن $(q = 1 - p)$. وجدير بالذكر بأنه كلما كان معامل صعوبة الفقرة موضع البحث لا يختلف جوهريا عن المعامل الافتراضي (0.50 مثلا)، كانت قيمة المعامل جيدة، أي أنها لا تختلف كثيرا عن المستوى المنشود.

تمرين: هل تختلف معامل صعوبة فقرة في اختبار تحصيلي التي كان مقدارها (0.70) عن معامل الصعوبة الافتراضي الذي كان مقداره (0.50) في ذات الإختبار والذي طبق على ذات المجموعة من الطلبة وعددهم (40) عند مستوى الدلالة (0.05).

الحل: يتم صياغة فرضيتي الدراسة على النحو التالي:

الفرضية الصفرية: لا تختلف معامل صعوبة الفقرة المحسوبة عن المعامل الافتراضي على مستوى دلالة (0.05). وبالرموز: $Ho : \pi = 0.50$

الفرضية البديلة: تختلف معامل صعوبة الفقرة المحسوبة عن المعامل الافتراضي على مستوى دلالة (0.05). وبالرموز: $H_A : \pi \neq 0.50$

ولأن مستوى الدلالة (0.05) والفرضية البديلة غير متجهة، فإن القيمة الحرجة للاختبار Z هي (±1.96). وعليه، يتم رفض الفرضية الصفرية إذا كانت القيمة المحسوبة لإختبار Z تزيد عن (+1.96) أو تقل عن (-1.96). ويتبع ذلك حساب قيم

ويتم التعويض في المعادلة لحساب قيمة الإختبار Z على النحو التالي:

$$Z = \frac{p - \pi}{\sqrt{\frac{pq}{n}}} = \frac{.70 - .50}{\sqrt{\frac{(.70)(.30)}{40}}} = 2.7$$

وعليه، يتم رفض الفرضية الصفرية وقبول البديلة، لأن القيمة المحسوبة للاختبار Z أكبر من القيمة المحددة مسبقا وهي (1.96). وهذا يشير إلى أن صعوبة الفقرة (0.70) تختلف عن معامل الصعوبة الافتراضي (0.50) اختلافا جوهريا.

معامل شدة العبارة

جدير بالذكر القول بأن معامل الصعوبة مؤشر كمي يستخدم في تقييم فقرات اختبارات التحصيل والذكاء والاستعدادات والقدرات المعرفية. أما عند الحديث عن قوائم الشخصية ومقاييس الاتجاهات والميول، نستخدم مفهوم "شدة العبارة" عوضا عن معامل الصعوبة. ذلك، لأن الاستجابة على عبارات تلك المقاييس تشير إلى مدى انطباق محتوى العبارة على الفرد من

ناحية، ودرجة انطباقها على الفرد من ناحية أخرى. إذ ليس من المعقول القول بأن عبارة المقياس صعبة على الفرد، بل يمكن القول فقط بأن تلك العبارة منطبقة أو لا، أو أنها تنطبق بشدة أو بدرجة متوسطة أو بدرجة ضعيفة. ويتم قياس شدة العبارة من خلال حساب تكرارها. ويتم تقييم اختلاف معامل الشدة باختلاف الفئة المستهدفة (ذكور، إناث)، (فائقو أو متوسطو التحصيل)، (مرتفعو، متوسطو، منخفضو الذكاء) وهكذا.

تمييز الفقرة Item Discrimination

إن الاعتماد على معامل الصعوبة للفقرة فقط وكمؤشر وحيد لا يفي بالغرض. لأنه لا يساعد في إجراء مقارنة بين أداء الممتحنين حسب مستوى أدائهم الكلي. لذلك فإن معامل تمييز الفقرة من الصفات الهامة في تحليل الفقرات. فتمييز الفقرة يساعد في تحديد قدرتها على التمييز بين الطلبة ذوي التحصيل المتدني والطلبة ذوي التحصيل العالي. فلو عرضنا اختبارا مقننا على مجموعة من الممتحنين، ووجدنا بالنهاية أن عدد الطلاب ذوي التحصيل المنخفض في مادة الإختبار مساويا عدد الطلبة من ذوي التحصيل المرتفع الذين أجابوا فقرة ما إجابة صحيحة، فإن هذه الفقرة لا تصلح لفحص الفروق بين الطلبة في التحصيل على هذا الإختبار. ولأن هذه الفقرة لا تميز أبدا بين مجموعتي الممتحنين، فهي غير جيدة، أو غير واضحة، أو على الأقل تحوي لبساً معيناً في متنها، أو في بدائلها، أو ربما في مفتاح تصحيحها.

ولتحديد الممتحنين من ذوي التحصيل المرتفع وذوي التحصيل المنخفض، لا بد من وجود محك Criterion، يتم التصنيف في ضوئه. فقد يتم استخدام درجات الممتحنين الكلية (مجموع العلامات على فقرات الإختبار) كمحك داخلي Internal Criterion، أما بالنسبة للاختبارات التي تقيس الميول والاستعدادات والمهارات، فيستخدم محكات خارجية External Criterion كمعدل تقديرات الطلبة نهاية العام الدراسي، أو تقديرات

المتدربين نهاية البرنامج التدريبي، أو حالة أفراد المهن المختلفة كمجموعات محكية.

تعددت طرق حساب معامل تمييز فقرات الإختبار وتنوعت، فبعضها يتوقف على درجات المجموعتين الطرفيتين Extreme Groups. ويعتمد الآخر على معاملات الارتباط Correlation Coefficients. وفيما يلي عرض لعدد من التمارين لحساب معامل تمييز الفقرة حسب آلية تصحيحها. إذ سيتم تناول معامل تمييز الفقرة الانتقائية ثم معامل تمييز الفقرة الصياغية (الإنشائية):

معامل تمييز الفقرة الانتقائية

تستخدم هذه الطريقة للفقرات التي تصحح ثنائيا (0 ، 1) فقط. ويتطلب حساب معامل التمييز بهذه الطريقة تحديد نقطتي قطع cut scores على المحك الداخلي المستخدم وهو الدرجة الكلية لعلامات الممتحنين. تلك النقطتين تستخدمان لإبراز مجموعة الأداء المرتفع ومجموعة الأداء المنخفض، وهذا يتطلب ابتداء ترتيب الممتحنين ترتيبا تنازليا حسب علاماتهم الكلية. ثم يتم تحديد أعلى 50% وأقل 50% (إذا كان عدد المفحوصين قليلا) أو أعلى 27 % وأقل 27 % (على اعتبار أن معامل تمييز الفقرة يكون حساسا، وأكثر استقرارا في حال استخدام هذه النسبة).

أشار كيللي Kelley (1939) إلى أن نتائج معامل تمييز الفقرة المحسوب على نسبة مئوية (27 %) لا تختلف كثيرا عن تلك المحسوبة على نسبة (30 %) أو (50 %) عندما يكون عدد المفحوصين كبيرا (أي بواقع 100 أو أكثر). أما كورتون (Cureton (1956، فينصح باستخدام النسب (29 %) أو (30 %) إذا لم يكن التوزيع إعتداليا. وبعد تحديد مجموعتي الأداء، يتم حساب معامل تمييز الفقرة D بطرح نسبة الذين أجابوها إجابة صحيحة من ذوي التحصيل المتدني (P_L) من نسبة الذين أجابوها إجابة صحيحة من مجموعة التحصيل العالي (P_u) كما في المعادلة التالية:

$$D = P_u - P_l$$

ويتراوح مقدار معامل التمييز بين (−1.00) و (+1.00) ، وكلما كان هذا المقدار المحسوب أكبر كلما دل على تميز أكبر للفقرة. فعندما يكـون تميز فقرة ما (+1) فهذا يعني أن جميع الطلبـة ذوي التحـصيل العـالي أجابوها بشكل صحيح، ولا أحداً من ذوي التحصيل المتدني أجابها.

كذلك، فإن هذه الفقرة ذات نوعية ممتازة، أما إذا كان معامل التمييز صفراً فهذا يعني أنها لم تفرق بين المجموعتين. وأسوأ تميز للفقرة عنـدما تكون قيمة D سالبة، حيث أن الطلبة ذوي التحصيل العالي لم يجيبوا على الفقرة إجابة صحيحة، بينما أجابها نسبة أكبر مـن الممتحنـين مـن ذوي التحصيل المتدني. وقد اقترح إيبل Ebel (1956) دليلا ما زال يستخدم فـي تصنيف الفقرات بناءا على قيم معاملات التمييز كما في الجدول التالي:

قيمة D	مدلول D
0.40 فأكثر	الفقرة مقبولة، وتميز جيـدا، وينـصح بإبقائهـا فـي الإختبار
0.30 – 0.39	الفقرة مقبولة، أو ربما تحتاج تعديلا طفيفا
0.20 – 0.29	الفقرة تحتاج تعديلا معقولا
أقل من 0.20	الفقرة تحذف أو يتم تعديلها جوهريا
سالبة	الفقرة تحذف ولا مانع من دراستها

وتتوقف عملية مراجعة الفقرات علـى نتـائج تحليـل فعاليـة المموهات التي سيرد شرحها لاحقا، لأن ذلك يساعد في الوقوف علـى كيفية أداء الممتحنين على كل بديل من بدائل الفقرة، ومن ثـم يتحـدد

البديل الذي ساهم في خفض معامل تمييز الفقرة.وعليه، يمكن حساب معامل التمييز للفقرات معيارية المرجع حسب الحالات التالية:

الحالة الأولى: عندما تتوفر بيانات إجمالية عن إجابة الفقرة

تمرين: ما معامل تمييز فقرة عرضت على(100) طالب، أجابها (20) من بين أفراد مجموعة الأداء المرتفع إجابة صحيحة، وأجابها (12) من بين أفراد مجموعة الأداء المنخفض إجابة صحيحة؟

الحل: في هذا التمرين، يكون عدد أفراد مجموعة الأداء المرتفع أو عدد أفراد مجموعة الأداء المنخفض يساوي (27). وعليه، يتم حساب معامل التمييز بإيجاد نسبة من أجاب الفقرة إجابة صحيحة من بين أفراد مجموعة الأداء المرتفع مطروحا منه نسبة من أجابها من أفراد المجموعة الدنيا كما في التالي:

$$D = P_u - P_l$$

$$D = \frac{20}{27} - \frac{12}{27} = \frac{8}{27} = 0.29$$

الحالة الثانية: عندما تتوفر بيانات تبين إجابات كل ممتحن على كل فقرة

تمرين: تم عرض علامات مجموعة ممتحنين على اختبار يتألف من 10 فقرات من نوع الاختيار من متعدد، وقد ظهرت درجات الممتحنين مرتبة تنازليا، مبينا أعلى وأقل 27 % منهم، وقد تم حساب معامل التمييز (D) للفقرات كما في السطر الأخير:

رقم المفحوص	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	الدرجة الكلية
01	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**0**	**1**	**1**	**9**
02	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**0**	**1**	**1**	**9**
03	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**0**	**1**	**0**	**8**
04	**1**	**0**	**0**	**1**	**1**	**1**	**1**	**0**	**1**	**1**	**7**
05	1	1	1	0	1	1	0	1	0	0	6
06	0	0	0	1	1	1	0	1	0	1	5
07	0	1	1	1	0	0	0	1	1	0	5
08	1	1	1	1	0	0	0	0	0	0	4
09	0	0	0	0	1	0	0	1	1	1	4
10	1	0	0	1	1	0	0	0	0	0	3
11	**1**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**3**
12	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**2**
13	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**1**
14	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**
D	**.75**	**0**	**.75**	**1**	**1**	**1**	**.75**	**-.25**	**1**	**.50**	

يلاحظ أن هناك فقرات معاملات تمييزها موجبة، وأخرى سالبة، وأربع فقرات تمييزها موجب وعالي جدا (1)، وفقرة واحدة معامل تمييزها صفرا. كما يجدر التذكير هنا أن حساب معامل التمييز بهذه الطريقة يتجاهل درجات المجموعة الوسطى والتي تشكل حوالي (46 %) من مجمل الممتحنين.

الحالة الثالثة: عندما تتوفر بيانات الإجابة على الفقرة على هيئة بطاقة

في هذه الحالة، تعطى بيانات الإجابة على الفقرة من قبل أفراد المجموعتين الطرفيتين (مجموعة الأداء المرتفع ومجموعة الأداء المنخفض). ومن ثم يتم تطبيق المعادلة:

$$D = \frac{T_u - T_l}{(n)(s)}$$

حيث يشير الرمزان T_u و T_l إلى مجموع درجات الممتحنين في مجموعة الأداء المرتفع ومجموعة الأداء المنخفض على الترتيب، بينما يشير الحرف n إلى عدد الأفراد في كل مجموعة من مجموعتي الأداء، ويشير الحرف S إلى مقدار العلامة المخصصة لتلك الفقرة. والتمرين التالي يبين هذه الحالة:

تمرين: يبين الجدول التالي توزيعا لأعداد إجابات 50 مفحوصا على بدائل الفقرة من نوع الاختيار من متعدد بواقع 4 بدائل. وإذا كان البديل (ب) هو الجواب الصحيح، أحسب معامل تمييز تلك الفقرة.

البديل	أعلى 27 % (14 مفحوصا)	أدنى 27 % (14 مفحوصا)
أ	صفر	2
ب*	10	3
ج	3	صفر
د	1	9

الحل: عند حساب معامل تمييز الفقرة المبينة كما في هذا الجدول، يجب التعرف أساسا على الجواب الصحيح، وهو هنا البديل (ب):

$$D=\frac{T_u-T_l}{(n)(s)}=\frac{10-3}{(14)(1)}=\frac{7}{14}=+0.50$$

وجدير بالذكر أن هذه الفقرة تتمتع بمعامل تمييز عالي وموجب. كما يجدر الإشارة إلى أنه يجوز لنا أن نقرب حجم أفراد مجموعة الأداء المرتفع أو المتدني لأقرب عدد صحيح. فقد تم تقريب (13.5) لتصبح (14).

الحالة الرابعة: معامل التمييز بطريقة معامل الارتباط

تتلخص الطريقة الإرتباطية لإيجاد معامل تمييز الفقرة بحساب معامل الارتباط بين الدرجة على الفقرة من جهة والدرجة الكلية على الإختبار من جهة أخرى. وعندما تشكل الدرجات على الفقرات متغيرا ثنائيا، بواقع (واحد أو صفر) والدرجات على الإختبار ككل متغيرا متصلا، فإن معامل الارتباط المحسوب هنا يدعى معامل الارتباط الثنائي المتسلسل Point Biserial Correlation حسب المعادلة التالية:

$$r_{pbis} = \frac{\mu_+ - \mu_x}{\sigma_x} \sqrt{\frac{p}{q}}$$

حيث يشير μ_x إلى متوسط الدرجات الكلية، ويرمز μ_+ إلى متوسط الدرجات الكلية للمفحوصين الذين أجابوا الفقرة إجابة صحيحة، σ_x إلى الإنحراف المعياري للدرجات الكلية، وتدل P على معامل الصعوبة للفقرة وتكون ($q = 1-P$).

ويشير معامل التمييز هذا إلى مدى اتساق الدرجات على كل فقرة مع الدرجات على الإختبار ككل. وعلى الرغم من وجود معادلة رياضية خاصة بحساب هذا المعامل كما ورد آنفا، إلا أن استخدام معادلة معامل ارتباط بيرسون بين متغيرين متصلين تؤدي إلى إعطاء نفس النتيجة. ومهما يكن من أمر، فإن مدلول معامل تمييز الفقرة لا يختلف باختلاف طريقة حسابه.

ولدى حساب معاملات التمييز لبعض الفقرات الواردة في التمرين السابق باستخدام الطريقة الإرتباطيه، تبين أن مجمل هذه المعاملات أقل من مثيلاتها التي حسبت باستخدام المجموعات الطرفية. فقد كانت معاملات التمييز للفقرات (q_1) و(q_2) و(q_3) هي (.59)، (.27)، (-.11) على الترتيب.

وبالرغم من ذلك كله، فإن المعامل المحسوب بهذه الطريقة ربما يضخم معامل التمييز الأصلي بسبب احتواء الدرجة الكلية للمفحوص على الفقرة المنوي حساب معامل تمييزها. ويمكن إهمال ذلك الخطأ في حالة العدد الكبير من فقرات الإختبار (25 فقرة أو أكثر).

أما عندما تكون درجة المحك (الدرجة الكلية) ثنائية بواقع ناجح وراسب مثلا، يكون معامل الارتباط المطلوب هو معامل الارتباط بايسيريال Biserial Correlation حسب المعادلة التالية:

$$r_{bis} = \frac{\mu_{+} - \mu_{x}}{\sigma_{x}} \left(\frac{p}{Y}\right)$$

حيث يشير μ_x إلى متوسط الدرجات الكليه، ويرمز μ_+ إلى متوسط الدرجات الكلية للمفحوصين الذين أجابوا الفقرة إجابة صحيحة ، σ_x إلى الإنحراف المعياري للدرجات الكلية ، وتدل P على معامل الصعوبة للفقرة، بينما يمثل الرمز Y مقدار مقطع المحور الصادي في التوزيع الطبيعي القياسي المقابل للقيمة الزائية التي تقابل معامل الصعوبة P كما هو في الملحق رقم (1) (Glass and Hopkins, 1996).

والمعروف أن معامل ارتباط فاي Φ يمكن أن يستخدم في حساب معامل الارتباط بين متغيرين ثنائيين إذا كان كلاهما ثنائي بحكم الطبيعة كالجنس (ذكر أو أنثى)، أو حالة المتقدم على اختبار قبول يستخدم للترشيح وظيفة ما (مقبول أو مرفوض). وجدير بالذكر أيضا أن قيمتي معامل فاي ومعامل بوينت بايسيريال متشابهتان لأنهما أساسا مشتقتان من معامل ارتباط بيرسون.

معامل تمييز الفقرة الصياغية (الإنشائية)

يتم استخدام طريقة المجموعتين الطرفيتين لتقدير معامل التمييز لهذا النوع من الفقرات وذلك بترتيب العلامات الكلية ترتيبا تنازليا، ثم يتم فرز المجموعتين بواقع أعلى 27 % كمجموعة الأداء المرتفع وأدنى 27 % كمجموعة الأداء المنخفض. ثم يتم تنظيم علامات الممتحنين على الفقرة المنوي حساب معامل تمييزها في المجموعتين، ثم تطبق المعادلة التالية:

$$D = \frac{T_u - T_l}{(n)(s)}$$

حيث يشير الرمزان T_u و T_l إلى مجموع درجات الممتحنين في مجموعة الأداء المرتفع ومجموعة الأداء المنخفض على الترتيب، بينما يشير الحرف n إلى عدد الأفراد في كل مجموعة من مجموعتي الأداء، ويشير الحرف S إلى عدد النقاط المخصصة للفقرة.

تمرين: يمثل التوزيع التالي علامات الممتحنين من المجموعتين العليا والدنيا على فقرة إنشائية وكان مخصصا لها 10 درجات. والمطلوب حساب معامل تمييزها:

	الممتحنون	علامات الممتحنين	مجموع العلامات (T)
مجموعة الأداء المرتفع	أحمد، سعيد، حمد، محمد، سيف، راشد، عبدالله، جاسم	10، 5، 8، 7، 5، 7، 9، 5	56
مجموعة الأداء المنخفض	فيصل، سالم، علي، طارق، صقر، طالب، سلطان، مبارك	4، 2، 9، 10، 3، 1، صفر، 1	30

الحل: نعوض القيم ($S = 10$)، ($n = 8$)، ($T_l = 30$)، ($T_u = 56$) في المعادلة كما يلي:

$$D = \frac{T_u - T_l}{(n)(s)} = \frac{56 - 30}{(8)(10)} = 0.325 \cong 0.33$$

وهذا معامل تمييز موجب وعالي حسب تصنيف إيبل.

دلالة معامل التمييز

مهما يكن من أمر فإن حساب معامل تمييز الفقرة بهذه الطريقة ما زال شائعا ومستخدما على نطاق واسع في الواقع الصفي وبطريقة يدوية. بينما يؤخذ عليها عدم وجود توزيع عيني Sampling Distribution معروف لقيمها. الأمر، الذي يجعل عملية اختبار دلالة اختلاف قيمة D عن الصفر، أو حتى المقارنة بين قيمتين من قيمها المحسوبتين أمرا غير ممكن (Crocker and Algina, 1986).

جاذبية (فعالية) المموهات Distracters Attractiveness

الأصل في المموه أن يكون جذابا للمفحوصين، وخاصة ممن ينتمون إلى مجموعة الأداء المنخفض. فإذا كان المموه يمثل إجابة خاطئة، فإن المفروض أن يختاره الطالب الضعيف. وبشكل عام، المفروض أن لا تقل نسبة الممتحنين الذين يختارون المموه عن 5 %، أكثرهم من فئة مجموعة الأداء المنخفض. ويمكن حساب معامل جاذبية المموه الذي يرمز له بالرمز DA بذات طريقة حساب معامل تمييز الفقرة حسب إحدى الصيغتين التاليتين:

$$DA = P_u - P_l = \frac{T_u - T_l}{ns}$$

والمموه الجيد والفعال هو ذلك البديل الذي يتمتع بمعامل جاذبية سالب وكبير. إذ يجب مراجعة أو تطوير أو استبدال أي بديل لا يختاره أحد من الممتحنين أو ذلك الذي تكون جاذبيته موجبة أو صفرا. وللتدرب على ذلك يمكن حساب معامل جاذبية المموهات (أ) و (ج) و(د) في الفقرة الواردة في

التمرين التالي. وهذا يقود إلى أن البـديلين (أ) و(د) جيـدان لأن معامـل جاذبية كل منهما سالبة، بينما ينصح بدراسة وتطوير البديل(ج).

تمرين: يبين الجدول التالي توزيعا لأعداد 50 مفحوصا على بدائل الفقرة من نوع الاختيار من متعدد بواقع 4 بدائل، وكان البديل (ب) هو الجواب الصحيح، أحسب جاذبية مموهات تلك الفقرة.

البديل	أعلى 27 % (14 مفحوصا)	أدنى 27 % (14 مفحوصا)
أ	صفر	2
ب*	10	3
ج	3	صفر
د	1	9

الحل: يتم حساب معاملات جاذبية المموهات:

$$DA_{أ} = \frac{0}{14} - \frac{2}{14} = \frac{-2}{14} = -0.14$$

$$DA_{ج} = \frac{3}{14} - \frac{0}{14} = \frac{3}{14} = 0.21$$

$$DA_{د} = \frac{1}{14} - \frac{9}{14} = \frac{-8}{14} = -.57$$

أما إذا اختار نسبة كبيرة من الممتحنين من أفـراد مجموعـة الأداء المرتفع أحد المموهات، فلا بأس من التعمق فيه، لأنه ربمـا يكـون هـو الإجابة الصحيحة لفقرة تم تصميم مفتاح التصحيح لها خطأ كما في التمرين التالي:

تمرين: يبين الجدول التالي توزيعا لاستجابات المفحوصين على بدائل فقرة وقد ظهر تاليا معاملات فعالية كل مموه من بدائل الفقرة:

البديل	أعلى 27 % (14 مفحوصا)	أدنى 27 % (14 مفحوصا)
أ	صفر	صفر
ب*	14	12
ج	صفر	2
د	صفر	صفر

$$DA_{أ} = \frac{0}{14} - \frac{0}{14} = \frac{0}{14} = 0.0$$

$$DA_{ج} = \frac{0}{14} - \frac{2}{14} = \frac{-2}{14} = -0.14$$

$$DA_{د} = \frac{0}{14} - \frac{0}{14} = \frac{0}{14} = 0.0$$

يتضح من ذلك أن هذه الفقرة تحوي بدائل ضعيفة جدا غير جذابة لغالبية الممتحنين من مجموعتي الأداء، كما في (أ) و (د). وينصح إما بدراسة البدائل من جديد، أو تطويرها على نحو آخر، أو تحويل الفقرة إلى فقرة من نوع الخطأ والصواب، أو من نوع الإكمال أو فقرة إنشائية. وجدير بالذكر أن تحليل إجابات الممتحنين على مموهات الفقرة يعتبر وسيلة للتعرف على مدى تحيز فقرات الإختبار item bias (Osterlind, 1983). وسيتم تناول موضوع تحيز الفقرات لاحقا نهاية هذا الفصل.

II. تحليل فقرات الإختبار محكي المرجع

عرفنا بأن الهدف الرئيس للاختبارات محكية المرجع CRT هو تقييم مستوى الأداء على مجموعة من المهارات التي تمثل مجالا سلوكيا ما. وعند تطوير هذا النوع من الإختبارات، يتجه الاهتمام إلى كل من طرق التدريس وكيفية تقديم المادة التي سيتم بناء اختبار لقياس مستوى الأداء فيها، وخطة تطوير الإختبار، ثم طبيعة الفقرات. إذ أن تصنيف الفقرات

حسب مقدار تباين كل منها أو الاهتمام بتباين الدرجات الكلية على الإختبار لا يعد هدفا رئيسا لهذا النوع من الإختبارات. ويعتمد تحليل فقـرات هـذا النوع من الإختبارات على المؤشرات الثلاثة التالية:

- صعوبة الفقرة Item Difficulty
- حساسية الفقرة Sensitivity
- مؤشر التوافق Index of Agreement

صعوبة الفقرة

لا يختلف حساب معامل الصعوبة فقرات اختبار CRT عنه في حالة فقرات اختبار NRT ، فهي نسبة الممتحنين الذين أجـابوا الفقـرة إجابـة صحيحة. ولكن التركيز على صعوبة الفقرة في هذا النوع من الإختبـارات لا يعتمد على مقدار ما تضيفه من تباين إلى الدرجات الكلية على الإختبار، كما أنها ليست مؤشرا أساسيا يستخدم لتقييم الفقرة.

وبشكل عام، يتوقع أن تكون الفقرات محكية المرجع أسهل عموما منه في حالة الإختبارات معيارية المرجع. أي أن متوسط معـاملات صـعوبة مجموعة من الفقرات محكية المرجع أعلى نسبيا مـن متوسـط صـعوبة الفقرات معيارية المرجع (عودة، 2000). فالفقرة التي معامـل صـعوبتها (0.50) تصنف على أنها متوسطة الـصعوبة فـي الإختبـارات معياريـة المرجع، وتكون صعبة جدا في الإختبار محكي المرجع شـريطة تطبيقهـا على نفس مجموعة الممتحنين. كما ينـصح بتحديـد متوسـط أو وسـيط معاملات صعوبة الفقرات المحكية التي تقيس هدفا ما، وتستخدم هذه القيمة في تقييم فعالية تدريس ذلك الهدف (Crocker and Algina, 1986).

وعندما تكون الفقرة قبيل التدريس سهلة جدا، تجعل مطور الإختبـار يتساءل ما إذا كان هناك ضرورة لتدريس هذا الجزء من المحتوى. أما في حالة الفقرات الصعبة بعد عملية التدريس، فهي تشير إلى أن التدريس غير

فعال أو أن مواصفات أو محتوى الفقرة لم يتم تغطيته بالهدف التعليمي. وربما يكون من الأهمية بمكان فحص التباين في معاملات صعوبة الفقرات التي تقيس هدفا واحدا.

من زاوية أخرى، عندما تختلف فقرة عن مجموعة الفقرات التي تنتمي لهدف ما بأن تكون غاية في الصعوبة أو السهولة، يجب فحصها من حيث: الصياغة الفنية، المنبهات غير المقصودة، خطأ مفتاح التصحيح، أو طول متن الفقرة أو احتوائها خطأ علميا معينا.

حساسية الفقرة Item Sensitivity

حساسية الفقرة لعملية التعليم أو معامل تمييزها. ويمكن تصنيف ذلك حسب طبيعة تصحيح الفقرة على النحو التالي:

I. معامل حساسية الفقرات المصححة ثنائيا (صفر، 1) ويقع في اثنين هما:

- **معامل كوكس وفارقاس**
- **معامل برينان**

اقترح كوكس وفارقاس (Cox and Vargas, 1966) طريقة لحساب معامل حساسية الفقرة محكية المرجع (S)، وهو قدرة الفقرة على التمييز بين الطلبة الذين تلقوا التدريس أو التدريب أو الإرشاد والذين لم يتلقوا. ولذلك، يتم مقارنة أداء هؤلاء الطلبة على اختبار تشخيصي قبل التعرض للتعلم وعلى الإختبار نفسه أو اختبار مكافيء له (أي استخدام اختبارين متوازيين كلاسيكيا أو عشوائيا) بعد التعلم حسب المعادلة التالية:

$$S = P_{post} - P_{pre}$$

حيث يشير كل من P_{pre} و P_{post} إلى نسبة الممتحنين الذين أجابوا الفقرة إجابة صحيحة قبل التدريس أو التدريب أو التوجيه وبعده على الترتيب. وتتراوح قيم S بين −1 و +1 وتفضل القيم الموجبة والكبيرة. وتكون الفقرة عديمة الحساسية للتدريس عندما تكون S صفرا، بمعنى أن

نسبة من أجابها إجابة صحيحة لا تختلف بوجود التدريس أو عدمه. ويمكن أن يطلق على هذا المؤشر معامل تمييز الفقرة محكية المرجع.

تمرين: طبق اختبار محكي المرجع (تشخيصي) يتألف من (35) فقرة على (10) طلاب قبل تعريضهم لبرنامج تعليمي علاجي وبعده. فـإذا تـم عرض استجابات هؤلاء الطلبة على أربع من تلك الفقرات كما في الجدول التالي. وكان المطلوب حساب معامل حساسية كل فقرة منها:

رقم الطالب	قبل التعليم (pre)				بعد التعليم (post)			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
1	0	0	1	1	1	0	1	0
2	0	1	1	1	1	1	1	0
3	0	1	1	1	1	1	1	0
4	0	1	1	1	1	1	1	0
5	0	0	1	0	1	1	1	0
6	0	0	1	0	1	1	1	0
7	0	0	1	1	1	1	1	0
8	0	0	1	1	1	0	1	0
9	0	1	1	1	1	0	1	0
10	0	1	1	1	1	1	1	1

الحل: يتم حساب معامل حساسية الفقرات:

$$S_{Q1} = P_{post} - P_{pre} = \frac{10}{10} - \frac{0}{10} = +1.0$$

$$S_{Q2} = P_{post} - P_{pre} = \frac{7}{10} - \frac{5}{10} = +0.2$$

$$S_{Q3} = P_{post} - P_{pre} = \frac{10}{10} - \frac{10}{10} = 0.0$$

$$S_{Q4} = P_{post} - P_{pre} = \frac{1}{10} - \frac{8}{10} = -0.7$$

يلاحظ أن الفقرات (Q1) و (Q2) تميز بشكل إيجابي، أي أن لديها حساسية موجبة تجاه برنامج التعليم العلاجي. بينما لم تبين الفقرة (Q3) أي حساسية، وأظهرت الفقرة (Q4) تمييزا سالبا. الأمر الذي يستدعي دراسة الفقرة (Q4) أو حذفها.

معامل برينان

اقترح برينان **Brennan (1972)** أسلوبا آخر لحساب معامل الحساسية وذلك بمقارنة مجموعتين من المفحوصين، أحدهما صنفت على أنها مجموعة متمكنة Master أو تلك المجموعة التي كانت درجات أفرادها على اختبار يقيس محتوى أو مجالا سلوكيا ما أعلى من درجة التمكن mastery cutoff score ، والأخرى غير متمكنة Non-master أو تلك التي تقع درجات أفرادها دون درجة التمكن. وعليه يتم حساب معامل برينان لحساسية الفقرة حسب المعادلة التالية:

$$S = \frac{U}{N_1} - \frac{L}{N_2}$$

ويشير كل من N_1 و N_2 ق درجات التمكن وتحتها على الترتيب.

ويشير الرمز U إلى عدد الذين أجابوا الفقرة إجابة صحيحة ممن درجاتهم الكلية فوق درجة القطع (التمكن)، بينما يرمز الحرف L إلى عدد الذين أجابوا الفقرة إجابة صحيحة ممن درجاتهم الكلية تقع تحت درجة القطع. وتتراوح قيم S بين −1 و +1 بتفضيل القيم الموجبة والعالية. ويبين الجدول التالي تخطيطا لمكونات المعادلة:

	إجاب الفقرة إجابة		المجموع
	إجابة صحيحة (1)	إجابة خاطئة (0)	
المتمكنون	U		N1
غير المتمكنين	L		N2

تمرين: أظهرت دراسة أن (70) طالبا تجاوزت درجاتهم الكلية درجة القطع (65) بينما بقي منهم (30) طالبا دون تلك الدرجة. فإذا أجــاب (40) طالبا من المجموعة المتمكنة على الفقرة إجابة صحيحة، وأجــاب عليهــا (20) طالبا من المجموعة الأخرى إجابة صحيحة كما في الجدول التــالي. فكيف نحسب حساسية الفقرة للتدريس؟

	أجاب الفقرة إجابة		المجموع
	إجابة صحيحة (1)	إجابة خاطئة (0)	
المتمكنون	40	30	70
غير المتمكنين	20	10	30

الحل: يتم التعويض في المعادلة والوصول إلى:

$$S = \frac{40}{70} - \frac{20}{30} = 0.57 - .66 = -0.10$$

تشير هذه القيمة إلى أن الفقرة ضعيفة الحساسية وسالبة.

II. معامل حساسية الفقرة غير المصححة ثنائيا أو الإنشائية

في حالة الفقرات محكية المرجع التي ربما تستخدم في الإختبــارات التشخيصية والتي تصحح بشكل متصل أو غير ثنائي (وقد يستخدم لــذلك

الغرض قاعدة تقدير rubricks مناسبة)، يمكن استخدام المعادلة لحساب معامل حساسيتها:

$$S = \frac{\overline{X}_{post} - \overline{X}_{pre}}{M - L}$$

حيث تشير الرموز $\overline{X}_{post}$ و $\overline{X}_{pre}$ إلى متوسط علامات المجموعة على الفقرة عندما يتم امتحانهم بعد انتهاء عملية التعليم وقبلها على الترتيب. ويشير الرمز M إلى الحد الأعلى للدرجة على الفقرة بينما يرمز الحرف L إلى الحد الأدنى للدرجة على الفقرة. أي أن العلامة على الفقرة تتراوح بين L و M.

تمرين: تأمل البيانات التالية التي تعرض استجابات عشرة من الطلبة قبل التعليم وبعده على فقرة، كانت الدرجة العظمى عليها (5)، علما بأن أقل علامة تعطى على هذه الفقرة (1). فكيف يتم حساب معامل حساسية تلك الفقرة:

علامات الطلبة على الفقرة		رقم الطالب
بعد التعليم post	قبل التعليم Pre	
5	2	1
5	3	2
5	1	3
4	5	4
4	5	5
4	4	6
5	4	7
4	2	8
5	2	9
3	1	10

يتم تطبيق المعادلة لحساب معامل الحساسية:

$$S = \frac{\overline{X}_{post} - \overline{X}_{pre}}{M - L} = \frac{4.4 - 2.9}{5 - 1} = \frac{1.5}{4} = 0.375 \cong 0.38$$

وهذا معامل حساسية موجب وعالي نسبيا.

تحيز الفقرة Item Bias

يعرف التحيز Bias بأنه خطأ منتظم systematic error في عملية القياس لصالح فئة معينة من المفحوصين، ويظهر هذا الخطأ في جميع مستويات الأداة بنفس القيمة والاتجاه، بحيث يؤدي إما إلى زيادة أو نقص في الدرجة على كل مستوى (Osterlind, 1983). وجدير بالذكر، فقد تم تناول مفهوم الخطأ المنتظم ومقارنته مع الأنواع الأخرى في الفصل الخامس من هذا الكتاب.

من ناحية أخرى، تكون الفقرة متحيزة عندما يتأثر أداء المفحوصين عليها بمتغيرات غير تلك التي قصدت في مضمون الفقرة، بمعنى أن الفقرة تكون متحيزة إذا كانت لا تتأثر بمصادر التباين عند جميع المجموعات والأفراد من نفس مستوى القدرة المقاسة (Crocker and Algina, 1986).

كذلك يرى ويلكوكس أن الفقرة تكون متحيزة عندما تعطي أفضلية لمجموعة من الأفراد دون غيرهم ليجيبوا عليها إجابة صحيحة. بمعنى أن الفقرات المتحيزة تؤدي وظائف مختلفة لدى المجموعات الفرعية المختلفة (Wilcox, 1985). وجدير بالذكر أن التحيز يظهر بسبب اختلاف اللغة أو العرق أو الثقافة أو الديانة أو جنس المفحوص.

أما الفقرة غير المتحيزة، فهي تلك التي تتساوى عليها احتمالات النجاح لجميع المفحوصين من ذوي القدرة نفسها، رغم اختلاف الأفراد في سمات أخرى غير تلك المقصودة بالقياس. وعليه، فإن الفقرة تكون متحيزة ضد مجموعة معينة، عندما يكون الأداء المتوقع لأفراد تلك المجموعة على الفقرات أقل من الأداء المتوقع للأفراد من مجموعات فرعية أخرى لهم نفس المستوى من القدرة على تلك الفقرة (Hills, 1989).

وباختصار، تكون الفقرة متحيزة عندما يختلف مستوى الأداء عليها تبعا لعوامل معينة لدى أفراد ينتمون لنفس مستوى القدرة العقلية. فقد يكون مستوى أداء فرد علي فقرة معينة أفضل من غيره من أفراد مجموعته الصفية أو العمرية بسبب أنه ينتمي لفئة إجتماعية معينة، أو جنس معين، أو تخصص معين، أو يقطن في حي معين، أو ينتمي لعرق معين. **وبشكل محدد، يمكن رصد مصادر تحيز الأداء على الفقرات في:**

- **التحيز العرقي:** ويظهر ذلك عندما يبرز اختلاف في مستوى الأداء على أداة القياس تبعا لانتماء الفرد إلى أقلية معينة في المجتمع. كان ينتمي إلى العرق السامي أو الآري أو غيرها. أو ربما كان من البيض أو من السود كما هو شائع في الولايات المتحدة الأمريكية وكندا وبعض البلاد الأوروبية.
- **التحيز للجنس:** كأن يظهر اختلاف في مستوى أداء الفرد بسبب كونه ذكرا أو أنثى.
- **التحيز الاجتماعي:** فقد يختلف مستوى أداء الفرد على فقرة معينة بسبب انتماء ذلك الفرد لمستوى إقتصادي وإجتماعي معين.
- **التحيز بسبب موقع السكن:** كما في المدينة أو الريف أو السكن في المناطق المحاذية للشواطىء، أو المناطق الصحراوية، أو تلك التي تقع على قمم الجبال، أو في مناطق حدودية، أو سياحية.
- **التحيز بسبب تخصص الفرد:** كأن ينشأ التحيز في الأداء على فقرة معينة بسبب اختلاف التخصص العلمي للمفحوص.

طرق الكشف عن تحيز الفقرة

تنوعت وتعددت طرق الكشف عن تحيز الفقرات. إذ يمكن فحص التحيز من خلال الطرق التالية:

- طريقة تحليل التباين الثنائي ANOVA. ويتم فحص وجود التحيز للفقرة إذا تم كشف وجود تفاعل ذي دلالة بين المجموعات الفرعية

والفقرات (كمتغيرين مستقلين) ومعامل صعوبة الفقرة (كمتغير تابع).

- تفحص قيم معاملات الصعوبة: وذلك بمقارنة قيم تلك المعاملات المحولة للفقرات على كل فئة من الفئات التي ينتمي إليها المفحوصون. وعليه تكون الفقرة متحيزة إذا ظهرت فروق ذات دلالة بين قيم صعوبة الفقرة حسب الفئة.
- تفحص معامل تمييز الفقره: كأن يتم مقارنة معامل الارتباط بين الأداء على الفقرة والدرجة الكلية على الإختبار لكل مجموعة من الأفراد ممن ينتمون إلى كل مستوى معين من مستويات المجموعة الفرعية.
- طريقة كاي تربيع: فقد تكون الفقرة غير متحيزة إذا لم تختلف إحتمالية الاجابة الصحيحة عليها من قبل المفحوصين لهم نفس مستوى القدرة وينتمون لمجموعات فرعية مختلفة. وكلما زاد مقدار كاي تربيع المحسوبة كلما زاد احتمال تحيز الفقرة.
- تحليل الاجابة على كل مموه من مموهات الفقرة، وهذا ينطبق على الفقرات ذات الاجابة المنتقاه عموما، أو الفقرات ذات الاختيار من متعدد بشكل محدد.
- تقييم منحنى خصائص الفقرة Item Characteristic Curve Method (ICC): وعليه تكون الفقرة غير متحيزة إذا تشابهت منحنيات خصائص الفقرة فيها للمجموعات الفرعية المختلفة، بمعنى أن الفقره تكون غير متحيزة إذا كانت معاملات صعوبتها وتمييزها وإحتمال الحصول على الاجابة الصحيحة بالتخمين لجميع المجموعات الفرعية (Crocker and Algina, 1986).

صدق الفقرة Item Validity

يعرف معامل صدق الفقرة (iv) بأنه حاصل ضرب الإنحراف المعياري لدرجات الممتحنين على الفقرة بمعامل ارتباط الدرجة على الفقرة بالدرجة الكلية على المحك الخارجي(y) (علامات الطلبة على اختبار له علاقة بالأداء على الفقرات موضع الاهتمام ، أو واقع الممتحنين على مهارات معينة أو برنامج تدريبي معين: راسب ، ناجح). وبالرموز فإن:

$$= s_i r_{iy} = \sqrt{p_i q_i}\, r_{iy}$$

حيث يرمز الحرف s_i إلى الإنحراف المعياري للدرجات على الفقرة، بينما يشير r_{iy} إلى معامل الارتباط بين الدرجات على تلك الفقرة من ناحية، ودرجات الممتحنين على المحك (Crocker and Algina, 1986).

ثبات الفقرة Item Reliability

يعرف معامل ثبات الفقرة (ir) بأنه حاصل ضرب الإنحراف المعياري لدرجات الممتحنين على الفقرة بمعامل ارتباط الدرجة على الفقرة بالدرجة الكلية على المحك الداخلي (x) وهو علامات الطلبة الكلية على الإختبار، وبالرموز فإن:

$$ir = s_i r_{ix} = \sqrt{p_i q_i}\, r_{ix}$$

ويرمز الحرف s_i إلى الإنحراف المعياري للدرجات على الفقرة، بينما يشير r_{ix} إلى معامل الارتباط بين الدرجات على تلك الفقرة من ناحية، ودرجات الممتحنين الكلية على الإختبار (Crocker and Algina, 1986). ومعاملا الارتباط (r_{ix} و r_{iy}) هما من نوع بوينت بايسيريال.

إن معرفة قيم ثبات وصدق فقرات الإختبار يفيد بدرجة كبيرة مطوري الإختبارات. إذ أن إضافة فقرة معينة يعمل على زيادة تباين الإختبار ككل. الأمر الذي يجعل مطور الإختبار يضيف فقرات حتى المستوى المقبول من تباين الإختبار ككل. حيث أن التباين الكلي للدرجات على الإختبار يساوي مربع مجموع معاملات ثبات فقرات ذلك الإختبار كما في المعادلة التالية:

$$S_x^2 = (\sum s_i r_{ix})^2$$

كما أن معامل صدق المحك للإختبار r_{xy} يساوي حاصل قسمة مجموع معاملات صدق فقرات الإختبار على مجموع معاملات ثبات الإختبار (Nunnally, 1978) كما في المعادلة التالية:

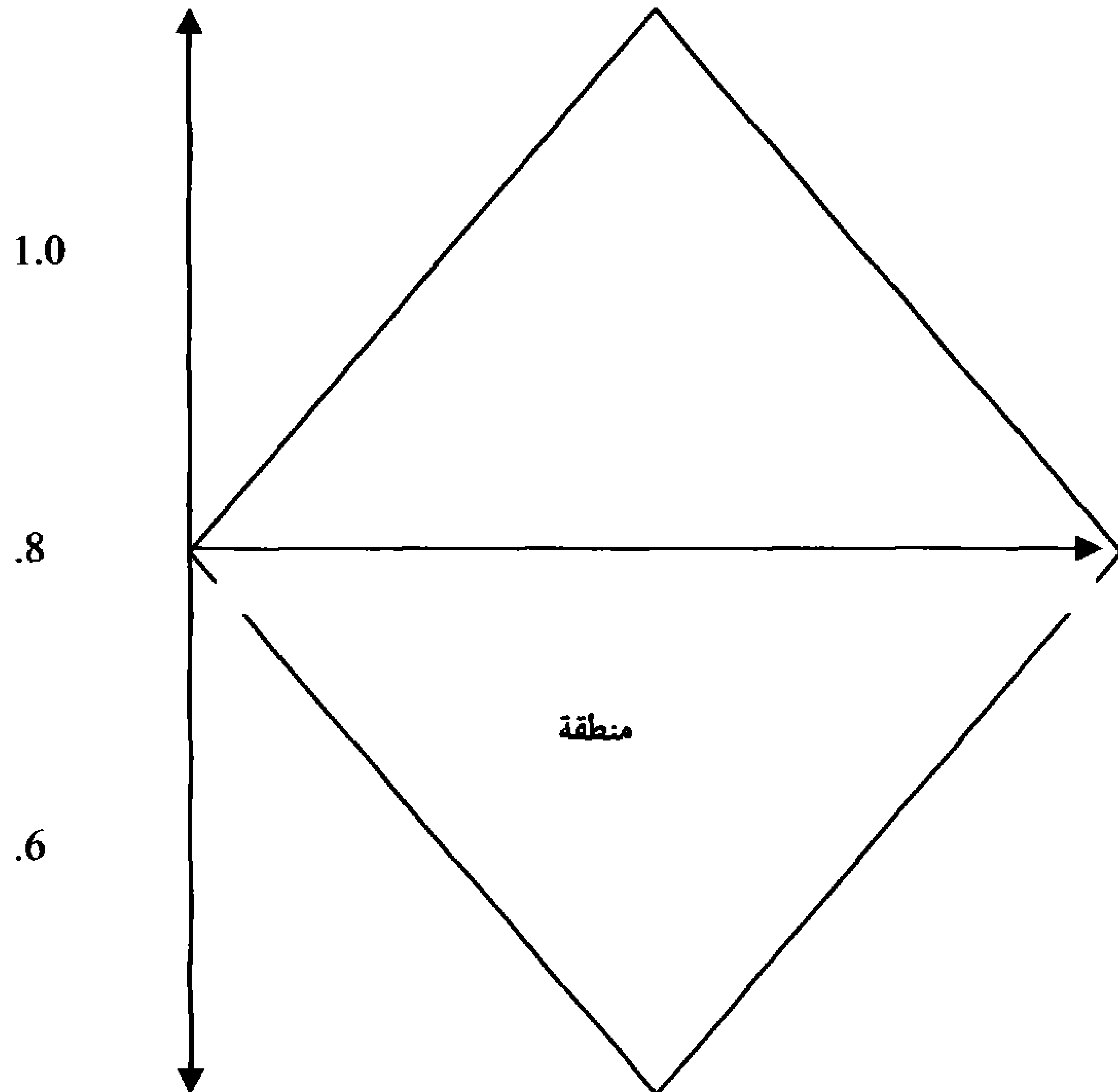
$$r_{xy} = \frac{\sum S_i r_{iy}}{\sum S_i r_{ix}}$$

تحليل فقرات الإختبار وحجم العينة

لا يوجد قانون مطلق أو قاعدة عامة تضبط الحد الأدنى لعدد الممتحنين من أجل إجراء دراسة تحليل الفقرات، بل هناك اتفاق على أن تحليل الفقرات يتطلب عددا كبيرا نسبيا من المفحوصين، وخاصة عندما يستخدم هذا التحليل في تطوير اختبارات مقننة وواسعة الاستخدام في أنحاء واسعة ، كاختبار (GRE) أو (TOEFL) وغيرها، مقارنة مع الإختبارات التي ربما يتم تطويرها لأغراض أكاديمية أثناء دراسة الدكتوراه أو لنيل درجة أكاديمية معينة (Crocker and Algina, 1986). وبشكل عام، يوصي نناللي (Nunnally, 1978) أن يتراوح عدد الممتحنين بين خمسة إلى عشرة أمثال عدد فقرات الإختبار كحد أدنى. أما استخدام نظرية السمة الكامنة IRT في تحليل الفقرات وبناء الإختبارات، فيتطلب عددا أكبر من الممتحنين لإجراء تحليل فقرات يمتاز بالثبات والموضوعية.

العلاقة بين معامل الصعوبة ومعامل التمييز

يعتمد تفسير كل من معامل الصعوبة P ومعامل التمييز D على الغرض من استخدام الإختبار وخصائص المجموعة المرجعية. كما يعتمد معامل التمييز على معامل صعوبة الفقرة. وعند الحديث عن العلاقة بين مؤشري فقرات الإختبار معياري المرجع ، نعرف أن تمييز الفقرة يعتمد على تباين أداء الممتحنين على تلك الفقرة. وتبلغ أقصى دلالات هذا التباين عند معامل صعوبة مقداره (.50) ويقل معامل تمييز الفقرة كلما ابتعدت قيم معامل الصعوبة عن (.50)، وتصل قيم D إلى أدنى قيمتها في حالة الفقرات الغاية في السهولة أو الغاية في الصعوبة. وبشكل عام، فإن الفقرات التي تتراوح معامل صعوبتها بين (.40) و (.60) يكون تمييزها مرتفعا كما في الشكل التالي الذي تم اقتباسه بتصرف من نيتكو (Nitko, 1983):



لا شك أن الفقرات ذات القدرة التمييزية العالية تؤدي إلى أن يكون معامل صعوبتها ضمن المدى الذي يحقق تباينا معقولا للفقرة، فقد قيل بأن مدى معاملات الصعوبة المحصور بين (.20) و (.80) بمتوسط مقداره (.50) يمكن أن يكون مقبولا(Nitko, 1998). أما بالنسبة للفقرات التي تقع خارج هذا المدى أي أقل من (.20) وأكبر من (.80) فقد تكون فقرات مناسبة من ناحية عملية على الرغم من أنها غير مناسبة من الناحية الإحصائية (عودة، 2000). فقد يلجأ واضع الإختبار إلى وضع فقرات سهلة لأغراض تشجيع الممتحنين أو اكتشاف قدراتهم المتقدمة على فقرات أصعب نسبيا.

من ناحية أخرى، تساعد البيانات المتعلقة بمعامل تمييز الفقرة في التعرف على الفقرات الضعيفة، بينما تستخدم بيانات معامل الصعوبة وتوزيع الاستجابات على بدائل الفقرة في اكتشاف مصدر الخلل في الفقرة التي تعاني من ضعف في القدرة التمييزية. وفي الجدول التالي الذي تم اقتباسه بتصرف من كروكر وألجاينا(Crocker and Algina, 1986)، تم عرض بيانات لاستجابات خمسين طالبا على اختبار صفي يتألف من خمس وثلاثين فقرة من نوع الاختيار من متعدد بواقع أربع بدائل لكل منها. تم الاكتفاء بالبيانات المتعلقة بالفقرات من (21) إلى (35). ويبدو أنه تم عرض نوعين من معاملات تمييز الفقرة هما الطريقة الطرفية (D) وطريقة بوينت بايسيريال (r_{pbis}) وكذلك درجة صعوبة كل فقرة، وعدد المستجيبين لكل بديل من بدائل كل من هذه الفقرات (a, b, c, d).

رقم الفقرة	نسبة الإجابات على بدائل الفقرة				حذف	P	D	r_{pbis}
	A	b	c	d				
21 −1	24	4	52	16+	4	.16	.00	-.06
22*	4	40	56+	0	0	.56	.67	0.48
23*	0	76+	12	12	0	.76	.50	0.45
24	4	28+	28	32	8	.28	-.17	-.12
25	16	12	0	72+	0	.72	-.17	-.29
26	0	4	52	44+	0	.44	.00	-.11
27	92+	0	8	0	0	.92	.33	.45
28*	8	68+	0	20	4	.68	.83	.61
29*	24	12	56+	8	0	.56	.50	.46
30	88+	0	0	8	4	.88	.17	.31
31	68+	12	4	16	0	.68	.17	.15
32*	20	20	8	52+	0	.52	1.00	.73
33	8	16	60+	16	0	.60	.00	.06
34*	20	20	8	52+	0	.52	.83	.59
35*	80+	0	0	4	16	.80	.50	.43

+: الجواب الصحيح ، *: فقرة ذات تمييز عال، _ :فقرة ذات تمييز منخفض

من هذا الجدول، يمكن الانتهاء إلى القرارات التالية:

- قبول الفقرات 22، 23، 28، 29، 32، 34، 35 بناءا على معيار إيبل المتمثل بقبول الفقرة التي معامل تمييزها يكون (0.40) أو أكثر.
- قبول الفقرات 27، 30 بناءا على قيم معاملات بوينت بايسريال للتمييز التي كانت (0.30) أو أكثر.
- الفقرات التي وضعت إشارة نجمة بجوار أرقامها تم قبولها على ضوء المعيارين السابقين. وهذه يمكن قبولها دون الحاجة إلى إجراء تعديلات.

- الفقرات التي تحتاج إلى تعديلات معينة، هي التي كان معامل تمييزها أقل من (0.20) على معايير إيبل. وقد تمت الإشارة إليها في الجدول بوضع خط تحت رقم كل منها وهي: 21، 24، 25، 26، 31، 33 .
- يلاحظ أن الفقرة رقم 30 ذات معامل تمييز مقبول حسب قيمة معامل بوينت بايسيريال وغير مقبول حسب معايير إيبل. وقد تم قبولها بحجة أن حساب معامل تمييزها أخذ بعين الاعتبار استجابات كافة المفحوصين وليس أفراد المجموعتين الطرفيتين فقط.
- الفقرة رقم 21، تبدو غريبة من بين الفقرات التي تنتمي لذات المجموعة. حيث أن معامل تمييزها سالب ومعامل صعوبتها (0.16). وتشير البيانات الواردة في الجدول إلى أن نسبة كبيرة من المفحوصين (52%) اختاروا البديل الخاطئ (c) بدلا من اختيارهم للجواب الصحيح (d) ذلك البديل الذي تم تصحيح الفقرة على أساسه. وبعد تفحص الفقرة والرجوع إلى المحتوى الذي تقيسه، تم الانتهاء إلى أن هذه الفقرة صححت خطأ، أي أنها فقرة بمفتاح خاطئ (Miskeyed Item).
- الفقرة رقم 24 ، كان تمييزها سالبا، وهي صعبة نسبيا لأن معامل صعوبتها (0.28)، وكان توزيع استجابات المفحوصين على ثلاثة من بدائلها (b, c, d) بشكل متساوي إلى حد ما. الأمر الذي أدى إلى الاستنتاج بأن الإجابة تمت على تلك الفقرة بشكل عشوائي (at random). وما زيادة عدد الذين حذفوها (8) إلا دليل على أن هناك احتمالية تقود إلى أن الفقرة كانت غامضة أو أن متنها يعاني من إسهاب ومبالغة. وهنا لا بد من العودة إلى

محتوى الفقرة وإعادة تعديله، أو حذف الفقرة أو إعـادة كتابتهـا بشكل معقول.

- الفقرتان: 25 و 26 ، يبدو أن كلا منهما يعاني من ضعف فـي بناء بعض البدائل التي لم يتم اختيار أي منها في الفقرتين كالمموه (c) في الفقرة رقم (25)، والمموه (a) في الفقرة رقم (26). فـي ضوء ذلك، بات من الضروري أن يتم استبدال هذين المموهين ببديلين أكثر منطقية وفعالية.
- الفقرتان: 31 و 33، كان معامل التمييز لكل منهمـا منخفـضا. وعلى الرغم من أن نتائج التحليل المبينة في الجدول لم تشر إلى مصدر الخلل في أي منهما، إلا أن فحص محتوى هاتين الفقرتين وفحص متضمن البدائل التي تم اختيارها من قبل المفحوصين من مجموعة الأداء المرتفع، ربما يساعد في الوصول إلـى استنتاج معين.
- بالرغم من أن الفقرة رقم 35 وهي آخر فقرة في الإختبار، كانـت تتمتع بمعامل تمييز عالي، إلا أن ارتفاع نسبة المفحوصـين (16 %) الذين حذفوها يشير إلى ضرورة تمديد وقت الإختبار قليلا لتمكين كامل المفحوصين من محاولة الإجابة عليها. هذا عـلاوة على أن البديلين (b) و (c) بحاجة إلى تعديل وتطوير جوهري.

نقاط للتأمل

- كلما كانت الفقرة صعبة جدا أو سهلة جدا، يصبح معامل تمييزيهـا منخفضا. هذا بالرغم من أننا نحتاج أن يحوي الإختبار مثل هـذه الفقرات لتحقيق التمثيل الأفضل للمحتوى الدراسي.
- يؤثر هدف الفقرة، كجزء من الإختبار ككل، على معامل تمييزها. ولا ننس أننا نتعامل مع درجات اختبار كليه، على أسـاس أنهـا

معيار داخلي عند تصنيف المفحوصين إلى مجموعتين طرفيتين. لذلك، فإذا أرادت مدرسة الفيزياء مثلا، قياس أهداف ومجالات محتوى مختلفة، مثل الحرارة، والضوء، والمغناطيسية، والكهربائية، فإنها تحتاج إلى أنواع متعددة من الفقرات. فإذا كانت هناك ما نسبته (5 %) فقط من المحتوى الدراسي، يتعامل مع القدرة على تطبيق المبادئ، وتم توزيع ذلك على مجالات المحتوى المختلفة، فان فقرات التطبيق هذه قد تكون ذات قوة تمييزية منخفضة، لان النسبة العظمى من فقرات الإختبار قد تهتم بقياس المعرفة والفهم. كما أن حذف الفقرات ذات القوة التمييزية المنخفضة يضعف أو يفسد صدق الإختبار. وطالما كانت المدرسة تقوم بالتدريس باتجاه تحقيق هدف التطبيق، فهي مجبرة على أن تبني فقرات تقيس القدرة على التطبيق. وبهذه الحالة، ولأن الإختبار الصفي النموذجي يقيس العديد من الأهداف التعليمية، نتوقع أن نجد أن "معاملات تمييز منخفضة، وإيجابية بنفس الوقت، وهذه هي القاعدة وليس الاستثناء" (Gronlund, 1981).

- مؤشرات تحليل الفقرة هي معلومات أولية (ليست مطلقة). يفترض بعض العاملين بالإختبارات خطأ بأن معاملات الصعوبة والتمييز ثابتة أو نهائية. حيث تتأثر معلومات تحليل الفقرة بطبيعة المجموعة الخاضعة للاختبار، أو عدد المفحوصين، والأسلوب التعليمي المتبع في تدريسهم، وأخطاء الصدفة، وربما موقع الفقرة في الإختبار. من ناحية أخرى، يجب أن يكون حكمنا على جودة الفقرة مبنيا على مدى قياسها أو عدم قياسها لهدف تعليمي ذي قيمة، وليس على حجم معاملات صعوبتها وتمييزها. وطالما كانت الفقرة تميز بشكل إيجابي، وكانت واضحة وغير غامضة، وطالما كانت خالية من العيوب، فإنه يجب الإبقاء عليها على الأقل لغرض الاستخدام المستقبلي الممكن.

- تجنب اختيار فقرات الإختبار على أساس خصائصها الإحصائية فقط. إذ أن أفضل طرق اختيار فقرات الإختبار، يعتمد على مبدأ أن تلك الفقرة التي تمتلك صعوبة وتمييز ملائمين للهدف المنشود من الإختبار. أما الفاعلية الإحصائية، فهي ليست شرطا إلزاميا، وخصوصا في الإختبارات التحصيلية من صنع المعلم في الصفوف الأولى، حيث يتم حساب إحصائيات الفقرة اعتمادا على (25-30) من المفحوصين. الأمر الذي يجعلها غير ثابتة نسبيا. ومن الأسباب الأخرى لعدم اختيار فقرات الإختبار على أساس خصائصها الإحصائية فقط - حتى ولو كانت العينة المستخدمة كبيرة نسبيا - هو احتمال تأثر صعوبة الفقرة بعامل التخمين، وبموقع الجواب الصحيح بين البدائل الأخرى (الاختيارات)، وبتسلسل الفقرة وموقعها بين فقرات الإختبار (Huck and Bowers, 1972). مع ذلك، بقي هنالك سبب آخر لتوخي الحذر، وهو أن الاختيار الإحصائي في بعض الحالات، يؤدي إلى اختبار غير معبر ويتسم بالانحياز.

في ضوء ذلك، وجد كوكس (Cox, 1964) أن الإختبار المؤلف من فقرات اختيرت عل أساس خصائصها الإحصائية فقط، قد يكون منحازا من حيث أن الفقرات لا تقوّم الأهداف التعليمية بنفس النسبة التي يكون عليها بنك الفقرات الأصلي، بمعنى أن درجة صدق المحتوى قد تتدنى فعلا. وأوصى كوكس باختيار الفقرات الأكثر تمييزاً ضمن كل فئة تصنيفية، وهذا أفضل من اختيار الفقرات الأكثر تمييزاً من بين بنك الفقرات برمته، حيث يكون البناء التصنيفي مهملا. وإذا لم يتم ذلك، فان المعلم لن يتمكن من التأكد من أن صيغة الإختبار النهائية ستقيس الأهداف التعليمية المحددة في جدول المواصفات في مرحلة التخطيط.

- توفر معلومات تحليل الفقرة خدمة مفيدة وقيمة في اختيار فقرات الإختبار الجيدة. لكن يجب استخدامها كمؤشر لتعريف الفقرات التي تتطلب تفحصا دقيقا، وليس كوسيلة للتخلص من الفقرات المشكوك فيها. يجب علينا أن لا نعزز ميلنا نحو اختيار الفقرات التي تمتلك مواصفات إحصائية معينة فقط. إذ يفضل استخدام الطرق العقلية كأساس للاختيار الابتدائي لفقرات الإختبار، ومن ثم تستخدم التقنيات الإحصائية الأخرى للتأكد من أحكامنا. وعلى أية حال، نعتقد انه حتى بعد تحليل الفقرة، يجب مراجعتها بعناية من أجل ضمان التمثيل الدقيق للمحتوى.
- إن اختيار عدد من الفقرات بناءا على نوعية أداء عينة من المفحوصين، قد لا يكون استخدامها بالضرورة فعالا لصالح عينة أخرى. لذا كان من الحيوي أن يتم تجريب الفقرات التي تم اختيارها على عينة جديدة ومستقلة من الأفراد للتأكد من فعاليتها، وتعرف هذه العملية بالصدق المتقاطع Cross Validation. ولتسهيل إجراءات هذه العملية، يلجأ أحيانا إلى تطبيق الفقرات التي يتم تطويرها على عينة كبيرة نسبيا من المفحوصين مرة واحدة (400 شخصا) مثلا. ويصار إلى إجراء عمليات تحليل الفقرات على نصف أفراد تلك العينة، ويتم انتقاء الفقرات الفعالة. ويجري تنفيذ دراسة تقاطع الصدق على النصف الآخر من أفراد العينة وعلى الفقرات التي تم انتقاؤها للتأكد من فعالية الفقرات.
- يمكن الكشف عن وجود التخمين في الإجابة على فقرة معينة، عندما تتساوى أعداد المفحوصين الذين أجابوا على بدائل الفقرة بصرف النظر عن كونهم ينتمون إلى مجموعة الأداء المرتفع أو مجموعة الأداء المنخفض كما في المثال التالي:

البديل	مجموعة الأداء المرتفع	مجموعة الأداء المنخفض
أ*	8	8
ب	4	5
ج	7	6
د	9	9

ويحدث هذا عندما تقيس الفقرة هدفا هامشيا لا يهتم به الطلبة، أو تمثل مستوى عقليا عاليا، أو أنها فقرة مشتقة من خارج المحتوى الدراسي الذي يمثله الإختبار. الأمر الذي يدفع نسبة منهم إلى التخمين العشوائي (عودة، 2000).

أسئلة وتمارين

س1: أجب بـــ "نعم" أو بـــ "لا" على العبارات التاليه:

*.إذا تساوت قيم المتوسط والوسيط والمنوال، يكون توزيع تلـــك القـــيم ملتو نحو اليمين.

*. تتوزع درجات طلبة الصف العاشر على فقرات في مستوى الصف الثامن توزيعا طبيعيا.

*. يستخدم الوسيط في التعبير عن النزعة المركزية لبيانات تحوي قيما شاذة.

*. المدى مقياس ثابت لتشتت مجموعة من الدرجات.

*. معامل الصعوبة المتوسطة للفقرة يجعل معامل تمييزها أكبـــر مـــا يمكن.

*. ارتفاع معامل صعوبة الفقرة يجعل معامل تمييزها سالبا.

*. لا تحتاج الفقرة التي معامل تمييزها موجب ويقع دون (0.20) إلـــى تعديل.

*. يعد المموه جيدا إذا كان معامل تمييزه موجبا وعاليا.

*. التباين هو الأكثر استخداما من بين مقاييس التشتت.

*. تعد مسألة الاهتمام بمقاييس تشتت الدرجات على الإختبارات محكية المرجع أكثر منها في حالة الإختبارات معيارية المرجع.

*. تعتبر صعوبة الفقرة أحد صفات الفقرة المطلقة، فهي صعبة أو سهلة في كل الظروف.

*. يختلف حساب معامل صعوبة فقرة الإختبار باختلاف كون الإختبار محكي أو معياري المرجع.

*. يختلف مفهوم صعوبة الفقرة كونها فقرة انشائية أو انتقائية.

*. تقع معاملات صعوبة الفقرات على مستوى القياس الاسمي.

س2: أكمل الفراغات في العبارات التالية:

*. تتراوح قيم معاملات تمييز الفقرة بين ---- و ----.

*. إذا تساوت قيم المتوسط والوسيط والمنوال لمجموعة من الدرجات، فإن شكل توزيعها هو ----.

*. تتراوح قيم معامل صعوبة الفقرة بين ---- و ----.

*. إذا كانت الدرجة الأكثر تكرارا في توزيع الدرجات أكبر من الوسيط، وهذا أكبر من المتوسط، فإن شكل توزيع هذه الدرجات هو ----.

*. إذا جلس مجموعة من طلاب الصف الخامس لامتحان من مستوى الصف العاشر، فإن درجات الطلبة ستتوزع توزيعا ----.

*. النقطة التي تقع في توزيع من الدرجات ومجموع انحرافات القيم عنها يساوي صفرا هي ----.

*. النقطة التي تقع في توزيع من الدرجات ويقع قبلها عدد من القيم يساوي تلك التي بعدها هي ----.

*. الجذر التربيعي لمعدل مربعات انحرافات القيم عن متوسطها هو----.

*. تقع معاملات صعوبة الفقرات على مستوى القياس ----.

*. مربع مجموع معاملات ثبات فقرات الإختبار يساوي ----.

*. حاصل ضرب الإنحراف المعياري لدرجات عدد من المفحوصين على فقرة معينة ومعامل ارتباط الدرجة على الفقرة بالدرجة على المحك الداخلي يعرف بـ -----.

*. إذا أجاب ربع المفحوصين على فقرة موضوعية إجابة صحيحة، فإن معامل صعوبتها يساوي -----.

*. إذا أخفق ثلاثة أرباع المفحوصين من الإجابة على فقرة موضوعية إجابة صحيحة، فإن معامل صعوبتها يساوي -----.

س3: يبين الجدول التالي توزيعا لأعداد 50 مفحوصا على بدائل الفقرة من نوع الاختيار من متعدد بواقع 4 بدائل، البديل (أ) هو الجواب الصحيح، أحسب معاملي تمييز وصعوبة تلك الفقرة. كيف تعلق على هذه الفقرة؟

البديل	أعلى 27 % (14 مفحوصا)	أدنى 27 % (14 مفحوصا)
أ*	9	3
ب	صفر	1
ج	4	صفر
د	1	10

س4: يمثل التوزيع التالي درجات المفحوصين من المجموعتين العليا والدنيا على فقرة إنشائية التي كان مخصصا لها 10 درجات والمطلوب حساب معامل صعوبتها:

	درجات الممتحنين
مجموعة الأداء المرتفع	10، 5، 8، 6، 9،5، 10، 5،
مجموعة الأداء المنخفض	4، 5، 4، 3، 6، 1، صفر، 1

س5: كم يصبح معامل صعوبة الفقرة التي أجابها 70 مفحوصا إجابة صحيحة من أصل 100 شخص حاول حلها، إذا تم تصحيحها لأثر التخمين العشوائي وكانت من نوع:

(1) الإختيار من متعدد بواقع أربعة بدائل،

(2) الصواب والخطأ

س6: أظهرت دراسة أن (80) طالبا تجاوزت درجاتهم الكلية درجة القطع (70) بينما بقي (20) طالبا دون تلك الدرجة. فإذا أجاب (50) طالبا من المجموعة الأولى على فقرة في اختبار محكي المرجع إجابة صحيحة، وأجاب عليها (30) طالبا من المجموعة الأخرى إجابة صحيحة. فكيف نحسب حساسية الفقرة للتدريس؟

س7: تأمل البيانات التالية وفيها استجابات عشر مفحوصين على فقرتين من أصل 12 فقرة من نوع الاختيار من متعدد. واحسب معامل الصعوبة لكل فقرة، ثم أوجد معامل تمييز كل منهما اعتمادا على أعلى 50 % وأقل 50%، ثم بناءا على أعلى 27 % وأقل 27 %. مبينا الفرق في النتائج.

المفحوص	الفقرة (1)	الفقرة (2)	الدرجة الكلية
1	0	1	12
2	0	1	15
3	1	1	16
4	0	0	10
5	1	1	7
6	1	0	5
7	1	0	6
8	0	1	10
9	1	1	15
10	1	0	13

س8: تأمل بيانات تحليل فقرات خمس من الفقرات الموضوعية كما في الجدول التالي ثم أجب عن الأسئلة التي تلي، مبينا أن إشارة (+) تشير إلى الجواب الصحيح:

رقم الفقرة	استجابات الطلبة على البدائل				حذف	P	D	r_{pbis}
	a	b	c	d				
21	11	5	**149 (+)**	11	0	0.84	0.15	0.23
22	34	11	127 (+)	4	0	0.72	0.30	0.27
23	11	4	329 (+)	128	1	0.18	-0.13	-0.13
24	30	34	1	111 (+)	0	0.63	0.64	0.52
25	6	0	20	150 (+)	0	0.85	0.34	0.34

أ) بالرغم من تساوي معاملي صعوبة الفقرتين 21 و 25 إلا أنهما تختلفان في معاملي تمييز كل منهما، كيف يمكن تفسير ذلك؟

ب) كيف يمكن تطوير الفقرة رقم (21) ؟

ت) أعط سببين وراء سالبية معامل تمييز الفقرة رقم (23)؟

الفصل الخامس

الخصائص السيكومترية لأدوات القياس

أولا: الثبات Reliability

- مقدمة
- تعريف الثبات
- النظرية الكلاسيكية في الثبات
  - الخطأ المعياري للقياس
  - التقدير الكمي للدرجة الحقيقية للمفحوص
- تقديرات الثبات
  - ثبات الاستقرار
  - ثبات التكافؤ
  - ثبات التكافؤ والاستقرار
  - مقاييس الاتساق الداخلي
  - طريقة هويت في تحليل التباين
  - ثبات الفرق بين درجتي اختبارين
  - ثبات التصحيح
- مقارنة بين طرق تقدير الثبات
- العوامل المؤثرة على الثبات
- ثبات الاختبارات محكية المرجع
- الثبات واستخدام الاختبار
- أسئلة وتمارين

5

المخرجات التعليمية

بعد دراسة هذا الفصل يتوقع من الطالب أن يكون قادرا على أن

- يميز مصادر تباين الخطأ في القياس.
- يقارن بين الدرجة الخام والدرجة الحقيقية والعلاقة بينهما.
- يستطيع تقدير العلامة الحقيقية بدرجة ثقة معينة.
- يحدد المفهوم النظري للثبات وعلاقته بمفهوم الخطأ المعياري للقياس.
- يحسب التقديرات المتنوعة للثبات ويفهم مدى اختلافها.
- يحدد العوامل التي تؤثر على تقدير معامل الثبات.
- يقارن بين طرق تقدير ثبات أدوات القياس والتقييم.
- يثمن أهمية الثبات في البيانات المستخدمة في صنع القرارات.

مقدمة

تعتمد دقة صنع القرار في أي مجال من المجالات الإنـسانية علـى نوعية البيانات ومدى تغطيتها للمتغيرات التي يتكون منها المجال موضـع البحث. وعادة ما تتكون هذه البيانات نتيجة الأداء علـى أدوات قيـاس أو أساليب تقييم معينة، مقننة كانت أو غير ذلك، وكذلك بناءا على المشاهدات الميدانية لموقف إنساني أو حدث اجتماعي ما، أومن خلال تقارير أصحاب العمل عن سلوك العاملين، أو في ضوء انطباعات أولياء الأمور في وصف السلوك اليومي لأبنائهم في أوضاع متكررة، تتمثل في كيفية سلوكهم فـي المنزل، أو درجة التفاعل مع أفراد الأسرة، أو نوعية اللعب مع الرفـاق، ومدى عمق وتنوع العلاقات التي يبنيها الطفل في المدرسة مع أبناء صفه.

وعند استخدام البيانات في صنع القرار، علينا أن نعرف شـيئاً عـن جودتها. إذ يجب أن تكون البيانات ثابتة، متناسقة، مستقرة، وموضـوعية، لأن ذلك يساعد في أن تكون الاستنتاجات التي نشتقها من تلـك البيانـات صادقة. وفي هذا الفصل، والفصل التالي، سـنتعرض لمفهـومي الثبـات والصدق. هذا بالرغم من أن مفهوم الصدق هو الأكثـر أهميـة، إلا أننـا سنناقش الثبات أولاً، لأن الصدق يتضمن الثبات إلى حد معين.

يشعر أغلب المشتغلين في القياس النفسي والتربوي أنه لـيس مـن الضروري للممارسين أن يفهموا الثبات بعمق كبير. وبالرغم من أن هـذا الفصل لا يقدم طريقة نظرية متقدمة، فقد يجده القارئ أكثر صـعوبة مـن الفصول الأخرى من الكتاب.

بالرغم من أن الثبات يشكل جانبا مهما في تحديد نوعية أداة القياس، إلا أنه صفة نسبية وموقفية من خصائص تلـك الأدوات. بمعنـى أن أداة القياس ثابتة في ضوء ظروف ومعطيات معينة، وتتغير هذه الصفة بتغيـر خصائص عينة المستهدفين ونوعية العوامل المحيطة بهم. وبالختام، فإنـه

من غير الممكن أن تجد مقياسا ثابتا تماماً ودائما علـــى مـــدى الظـــروف والأزمان (Mehrens and Lehman, 1991).

إن اتباع إجراءات بناء الاختبارات التي تم تناولهـــا فـــي الفـــصول السابقة تؤدي بدرجة كبيرة إلى التوصل إلى اختبار ثابت. وسيتم التعرض إلى عدد من العوامل التي يتوقع أن تؤثر في معامل ثبـــات أدوات القيـــاس عموما.

وجدير بالذكر أن لا يتطلب الفهم العميق لهذا الفصل أي قدرة حسابية خارج نطاق ما هو مطلوب في الفصول السابقة. ولهؤلاء القراء الذين لديهم خلفية كمية في الإحصاء والرياضيات، ويفضلون معالجـــة أكثـــر عمقـــاً للموضوعات، نوصي بضرورة الاطلاع على المصادر التالية: (Crocker and Algina, 1986; Cronbach, 1971; Feldt & Brennan, 1989; Nunnally, 1978; Stanley, 1981)

تعريف الثبات

يمكن تعريف الثبات على أنه درجة الاتساق أو التجانس بين نتـــائج مقياسين في تقدير صفة أو سلوك ما. وفي ضوء ذلك، يتوقـــع أن تكـــون درجات الفرد ثابتة، إذا كانت متشابهة تحت ظروف قياس قليلة الاختلاف. فمثلا، إذا قمنا بقياس وزن شخص ما مرتين، فإننا نأمل أن نحصل علـــى نفس القياس تقريباً إذا استخدمنا مقياساً مختلفاً أو قمنا بوزن الفرد بعد يـــوم واحد فقط. وإذا قمنا بقياس مستوى تحصيل شخص ما، نأمـــل أن تكـــون القياسات متشابهة ولكن غير متطابقة تماما تحت إجراءات مختلفة أو على أيدي مديرين مختلفين باستخدام مصححين مختلفين، أو خلال أوقات مختلفة من يوم واحد. وبعبارة أخرى، نأمل أن نعمم من الدرجة الخاصة التي يـــتم الحصول عليها إلى الدرجة التي قد يمكن الحصول عليهـــا إذا مـــا كانـــت الظروف مختلفة قليلاً جداً.

تتفاوت درجة ثبات أداة القياس حسب المجال الذي تعمل به، فالمقاييس التي تهتم بقياس الخصائص الجسمية تتمتع بثبات أفضل نسبيا من أدوات قياس السمات العقلية والوجدانية. وهذا يعود إلى الأسباب التالية:

- تقاس الخصائص الجسمية مباشرة.
- تتمتع أدوات القياس المستخدمة في القياس الجسمي بدقة كبيرة نسبيا.
- تتمتع الخصائص الجسمية بثبات نسبي عموما.

وبالرغم مما سبق، **فقد يظهر هناك بعض من عدم الثبات أو عدم الاتساق الداخلي حتى في عمليات القياس الجسمي.** فإذا أردنا دراسة ثبات قياس وزن شخص ما، فإننا نستطيع أن نقوم بذلك بطرق عديدة. فقد نجعل الفرد يصعد وينزل من على الميزان، ثم نسجل وزنه في كل مرة، وقد تختلف هذه الأوزان المسجلة، لأن الفرد قد يقف بأوضاع مختلفة على الميزان بين مرة وأخرى. وربما لا يقرأ الشخص القائم على القياس مؤشر الميزان بشكل صحيح تماما، أو أنه لا يسجل القراءة بالشكل الصحيح أيضا.

ومن الطرق الأخرى لفحص الاتساق أو التجانس الداخلي والتي نستطيع بها قياس الوزن، هي أن نسجل وزن الشخص كما حصلنا عليه من عشرة مقاييس مختلفة ونقارن بين هذه القيم. وقد تتباين القيم للأسباب المذكورة آنفاً، كما قد تتباين بسبب أي اختلاف كان في المقاييس. لذا، يتوقع أن يحصل الفرد على مجموعة من القيم أكثر تباينا أو أقل اتساقا.

من هنا، ومن خلال استعراض الطرق التي سبقت، يمكن أن نحصل على معلومات حول الاتساق من خلال تحديد كمية التباين الموجود في درجات الفرد الواحد الذي يعرف **بالتباين ضمن الفرد** Intra-individual variability. وهذا التباين يسمى عادة بالخطأ المعياري للقياس.

كما يمكن دراسة الاتساق عندما يكون لدينا مجموعة من الناس يزنون أنفسهم مرتين (بتغيير المقاييس، أو تغيير عدد المرات، أو تكليف أكثر من قارئ أو من بسجل القياس). ويحددون بعد ذلك ما إذا كانت الأوزان النسبية للأشخاص قد بقيت كما هي. ويعطينا هذا تقديراً لثبات المقياس، ويعرف هذا التباين بين القياسات **بالتباين بين الأفراد** Inter-individual variability.

في ضوء ما تقدم، يبدو من غير الواقعي في القياس النفسي أن يتم قياس الفرد الواحد عدة مرات، إذ من الطبيعي أن لا نحصل على أي مقياس مباشر للتباين ضمن الفرد. وعلى أية حال، تقدم نظرية الثبات طرقاً متنوعة لتقدير هذا التباين من خلال قراءة بيانات التباين بين الأفراد، كما سنرى لاحقا.

ولا يخفى أن هناك أساليب متعددة تختلف في مدى تقدير ثبات القياس، كل واحد منها يسمح لمجموعة معينة من مصادر التباين بالتأثير على القيم المستخرجة على النحو التالي:

- قد يتغير مقدار الصفة أو السمة المراد قياسها من وقت إلى آخر (**خطأ عدم استقرار السمة: Trait instability**)؛
- قد يؤثر عدد ونوع الأسئلة التي يتم إعدادها لقياس مستوى الفرد المعرفي أو المهاري أو الوجداني على درجته. أي أن التساؤل هنا يتعلق بمدى تغطية الفقرات للمستويات المنوي قياسها (**خطأ المعاينة**: Sampling Errors)؛
- أي تغيير في الإرشادات، أو في الضبط الزمني للأداء على مكونات المقياس، أو في قراءة التعليمات (استخدام اللغة الفصحى أو اللهجة التي يفهمها المستجيب)، أو الألفة مع إدارة الاختبار يمكن أن يؤدي

إلى تباين في درجات الممتحنين (**خطأ إداري**: Administrative Errors)؛

- عدم الدقة في تصحيح ورقة الإجابة من حيث ضعف قراءة مفتاح التصحيح، أو احتساب درجات على فقرات معينة تندرج تحت بعد معين في المقياس وهي في الواقع تنتمي إلى بعد آخر. وهذا سيؤثر فعلا على درجات المفحوصين (**خطأ تصحيح**: Scoring Error)؛
- إن عوامل صحية (جسمية أو نفسية) مثل الدافعية، والتركيز، ودرجة التعب، والقدرة على التخمين يمكن أن تكون مصدرا لإحداث تباين في درجة المفحوص (**خطأ شخصي**: Personal Error).

ويمكن تصنيف الأخطاء حسب طريقة تأثيرها إلى:

(1) **أخطاء منتظمة** Systematic Errors ، وهي الأخطاء التي يكون تأثيرها في درجات المفحوصين متشابها من حيث المقدار والاتجاه. كأن يعطي الميزان مثلا زيادة أو نقصا بمقدار (3 كيلوغرامات) في وزن كل شخص. أو هي مقدار الإضافة التي يعطيها معلم الصف إلى درجة كل طالب في الصف (درجتين مثلا). هذا النوع من الأخطاء يسهل اكتشافه وبالتالي يمكن ضبطه. من ناحية أخرى، **يؤثر الخطأ المنتظم في صدق الاختبار وليس في ثباته.**

(2) **أخطاء عشوائية** Random Errors ، وهي الأخطاء التي تؤثر في درجات المفحوصين بمقادير واتجاهات مختلفة، وهي ترجع إلى عوامل الصدفة. فقد تعمل على زيادة درجات البعض، أو تؤدي إلى إنقاص بعضها، وقد لا تؤثر أبدا على بعض آخر من

الدرجات. فهي تشبه مدى تأثير ضوضاء أو عزف موسيقي أو مناقشات خارج قاعة الامتحان على واقع الطلبة الذين يجلسون لأداء امتحان ما. هذه الأخطاء يصعب التحكم بها أو الحد من تأثيرها. لذا، **يؤثر الخطأ العشوائي في ثبات الاختبار وفي صدقه.**

يسمى التباين في درجات الفرد بـ تباين الخطأ، وتعرف مصادر التباين (مثل عدم استقرار الصفة أو خطأ المعاينة) بـ مصادر الخطأ. وكلما كانت الأخطاء أقل، كلما كان القياس أكثر اتساقا (ثباتاً). وبهذه الخلفية العامة، نرى أنه لا بد من مناقشة موجزة للنظرية الكلاسيكية Classical Test Theory للثبات.

النظرية الكلاسيكية في الثبات

يمكن توضيح النظرية الكلاسيكية للثبات بأفضل وجه من خلال البدء بالدرجات الملاحظة (X) وهذه الأخيرة هي الدرجات التي يحصل عليها الأفراد على أداة القياس. وهذه الدرجات الملاحظة تشتمل على أجزاء مكونات عديدة. وفي أبسط صورها تتألف من درجة حقيقية (T) ودرجة خطأ (E)، فيكون:

$$X = T + E$$

وتشبه الدرجة الحقيقية إلى حد ما، ما يشير إليه علماء النفس بالدرجة الكونية Universal Score (Cronbach and Others, 1972). وهي الجزء من الدرجة الملاحظة غير المتأثر بالخطأ العشوائي. ومعروف كذلك أنه من غير الممكن ملاحظة الدرجة الحقيقية. وجدير بالذكر أن أي خطأ منتظم (كما في مقدار الإضافة (رطلين) التي يظهرها ميزان على وزن كل فرد)، لا يؤثر على الثبات أو الاتساق. وبذلك، يعد هذا الخطأ جزءا مهما، أو هو الجزء غير المتغير من درجة الفرد الملاحظة. وفي الواقع، قد لا يكون الافتراض بأن الدرجة الحقيقية للشخص ثابتة أو غير متغيرة،

افتراضاً دقيقاً. مع ذلك، فإن النظرية الكلاسيكية، تعتبر التغيرات في درجة الفرد الملاحظة ناجمة عن الخطأ.

ومما لاشك فيه أن الناس يختلفون عن بعضهم البعض فيما يتعلق بدرجاتهم الحقيقية ودرجاتهم الملاحظة. ولأن الأخطاء يفترض أن تكون عشوائية، فإن الأخطاء الإيجابية والسلبية يلغي أحدها الآخر، وبالتالي سيكون متوسطها صفرا. من ناحية أخرى، لا ترتبط الأخطاء العشوائية مع الدرجات الحقيقية ($r_{te}=0$) أو مع بعضها البعض ($r_{e1e2} = 0$). وبأخذ هذه الافتراضات بعين الاعتبار، نستطيع أن نكتب التباين في الدرجات على الاختبار كما يلي:

$$S_x^2 = S_t^2 + S_e^2$$

حيث تشير S_x^2 إلى تباين الدرجات الملاحظة لمجموعة من الأفراد، بينما تشير S_t^2 إلى تباين درجاتهم الحقيقية، وتعبر S_e^2 عن تباين الخطأ لتلك الدرجات. ويعرف معامل الثبات ($r_{xx'}$) على أنه نسبة التباين في الدرجة الحقيقية إلى التباين في الدرجة الملاحظة كما في المعادلة التالية:

$$r_{xx'} = \frac{S_t^2}{S_x^2}$$

فالثبات يخبرنا إلى أي مدى يعزى التباين الملاحظ إلى التباين الحقيقي. والرمز (r_{xx}) يستخدم للثبات، وتتراوح قيمة هذا المعامل بين الصفر والواحد. ويمكن توضيح علاقة مقدار تباين الدرجة الحقيقية وتباين الخطأ ومعامل الثبات كما في الشكل (1) المقتبس من براون (Brown, 1976):

معامل ثبات منخفض

تباين الدرجات الحقيقية S_t^2	تباين الخطأ S_e^2
تباين الدرجات الملاحظة S_x^2	

معامل ثبات مرتفع

تباين الدرجات الحقيقية S_t^2	تباين الخطأ S_e^2
تباين الدرجات الملاحظة S_x^2	

نموذج يبين علاقة تباين أجزاء الدرجة ومعامل الثبات

والمعادلتان السابقتان، هما اللتان نشتق منهما معظم الصيغ الشائعة المتعلقة بالثبات والخطأ المعياري للقياس. فإذا أعدنا خلط المعادلات السابقة، فإننا سنحصل على المعادلة التالية التي كثيرا ما يعبر بصيغتها عن معامل الثبات:

$$r_{xx} = 1 - \frac{S_e^2}{S_x^2}$$

من ناحية أخرى، يعرف مؤشر الثبات reliability index (r_{tx}) بأنه الجذر التربيعي لمعامل الثبات كما في المعادلة التالية:

$$r_{tx} = \sqrt{r_{xx'}} = \sqrt{\frac{S_t^2}{S_x^2}} = \frac{S_t}{S_x}$$

الخطأ المعياري للقياس Standard Error of Measurement

من خلال حل المعادلات السابقة، يتم حساب (S_e) حسب المعادلة التالية:

$$S_e = S_x\sqrt{1 - r_{xx'}}$$

ويدعى ذلك بالخطأ المعياري للقياس. وهو عبارة عن تقدير مقياس القدرة التباينية التي ذكرت سابقاً. ولأننا لا نستطيع عادة أن نقيس الشخص

باستمرار. وكثيراً ما يفهم من الانحراف المعياري للـــدرجات الملاحظـــة لشخص معين (من خلال إعادة إجراء الاختبار ذاته لمرات عديـــدة) شـــيئاً حول الدرجة الحقيقية للفرد في ذلك الاختبار. ومن التعريف، فإن الدرجـــة الحقيقية، التي رمز لها بالحرف (T) للفرد لا تتغير. وإذا ما أعدنا الاختبار على نفس الشخص لعدة مرات، سيكون هناك شيئاً من عدم الاتساق (خطأ)، وعليه فإن الدرجة الملاحظة(X) لهذا الشخص الواحد ستتباين، وقد تكون أكبر من درجة T، أو أقل منها أحيانا. وإذا ما افترضنا بأن الأخطاء ضمن درجات الفرد – خلال الجلسات الاختبارية – هي أخطاء عـــشوائية، فـــإن الأخطاء الإيجابية والسلبية ستلغي إحداها الأخرى، وسيصبح متوسط الخطأ صفراً. لذا فإن متوسط الدرجات الملاحظة عبر الاختبارات المتكررة هـــي الدرجة الحقيقية للفرد $(\overline{X}_i = \overline{T}_i)$، إذ أن رمز (i) يعني الفرد.

التقدير الكمي للدرجة الحقيقية للمفحوص

من المفروض أن تقع درجات الفرد الملاحظة ضمن توزيع اعتدالي حـــول درجته الحقيقية. يكون الانحراف المعياري لتلك الدرجات عبر الاختبارات المتكررة واضحاً إذا ما تفحصنا المعادلة التالية:

$$S_x^2 = S_t^2 + S_e^2$$

ولأجل الإعادة، يفهم الخطأ المعياري للقياس على أنه توفير معلومات حول تباين درجات المفحوص في اختبارات متكررة. إذ مـــن الطبيعـــي أننـــا لا نعطي للفرد نفس الاختبار لمرات عديدة، لأن ذلك شئ غير اقتصادي ولأن هذه الاختبارات المتكررة يمكن أن تؤدي إلى تغييرات جسمية أو نفسية في

الفرد (كالتعب والقلق). لذا، فإنه من المعتاد أن يتم تقدير الخطأ المعياري للقياس بناءا على بيانات المجموعة التي تأخذ الامتحان.

بناءا علي ما تقدم، يفهم بأننا يمكن أن نفسر درجة كل شخص باستخدام الخطأ المعياري الذي تم تقديره بناءا علي بيانات المجموعة. علي الرغم من أن هذا التفسير ربما يقود إلي نوع من سوء الفهم، خاصة إذا كانت المجموعة غير متجانسة.

ومهما يكن من أمر، فإن الخطأ المعياري للقياس يتمتع بامتياز تفسيري على معامل الثبات، حيث أنه يسمح لنا بأن نذكر مدى تباين درجة الفرد حسب اعتقادنا. وكثيراً ما يستخدم الخطأ المعياري للقياس لما يسمى بتفسير الفئة Band Interpretation. وتفسير الفئة هذا يساعد في تغطية فكرة عدم دقة القياس. (نستطيع أن نفكر بالفئة على أنها مدى خطأ الدرجة). فإذا افترضنا أن الأخطاء عشوائية، فإن الأخطاء الملاحظة لدرجات فرد معين ستوزع بشكل طبيعي حول الدرجة الحقيقة في اختبارات متكررة. لذا، فإننا نستطيع القول بأن الدرجات الملاحظة لشخص معين ستقع بين تباين الخطأ الموجب والسالب ($\pm 1S_e$) لدرجته الحقيقية بنسبة 68 % تقريباً من الوقت، أو بين ($\pm$ 1.96 S_e) من درجته الحقيقية حوالي 95 % من الوقت. وبالطبع، نحن لا نعرف الدرجة الحقيقية، لكن يستطيع الفرد أن يستنتج وبنسبة تأكد مقدارها (68 %) أو (95 %) أو (99 %) أن درجة الفرد الحقيقية هي ما بين ($\pm S_e$) أو ($\pm 1.96S_e$) أو ($\pm 2.58S_e$) لدرجته الملاحظة على الترتيب.

ويعرض الجدول التالي الذي اقتبس بتصرف من شيس كلنتون (1978) Chase Clinton قيما للخطأ المعياري للقياس بدلالة قيم الانحراف المعياري لمجموعات من الدرجات ومعاملات ثباتها:

قيم الخطأ المعياري للقياس بدلالة معامل الثبات والانحراف المعياري

الانحراف المعياري	معامل الثبات $r_{xx'}$		
S_x	.90	.80	.70
4	1.3	1.8	2.2
5	1.6	2.2	2.7
6	1.9	2.7	3.3
7	2.2	3.1	3.8
8	2.5	3.6	4.4
9	2.8	4.0	4.9
10	3.2	4.5	5.5
20	6.3	8.9	11.0
24	7.6	10.7	13.1

يلاحظ أن قيم الخطأ المعياري للقياس تزداد كلما قل معامل الثبات من ناحية، وتثبيت قيمة الانحراف المعياري من ناحية أخرى. كما أن قيم الخطأ المعياري للقياس تزداد بزيادة قيم الانحراف المعياري مع تثبيت قيمة معامل الثبات.

مثال: إذا كان معامل الثبات لإختبار الإستعداد الدراسي ($r_{xx'}$) يساوي (0.91) والانحراف المعياري للدرجات عليه يساوي (15)، فما مقدار الخطأ المعياري لقياس تلك الدرجات؟ وإذا حصلت سعاد على درجة مقدارها (112)، فكيف يمكن تقدير درجتها الحقيقية بنسبة ثقة (68 %) و (95 %) و (99 %) ؟

$$S_e = S_x \sqrt{1 - r_{xx'}} = 15 \sqrt{1 - 0.91} = 4.5$$

وعند تفسير المدى للخطأ المعياري في القياس للدرجات الملاحظة في أعلاه، سيكون كما يلي: إذا حصلت سعاد على درجة مقدارها 112، فإننا نستطيع أن نكون واثقين:

1. بنسبة 68 % من أن درجتها الحقيقية تقـع بـين (112 ± 4.5) أي أنها تتراوح بين (107.5) و (116.5).
2. بنسبة 95 % من أن درجتها الحقيقة تقع بين ((112±1.96(4.5)) أي أنها تتراوح بين (103.18) و (120.82)
3. بنسبة 99 % من أن درجتها الحقيقة تقع بين ((112±2.58(4.5)) أي أنها تتراوح بين (100.39) و (123.61).

كما أنه من الممكن تقدير الدرجة الحقيقية بدلالة معامل الثبـات للإختبـار المشتقة منه تلك الدرجة باستخدام المعادلة التالية. تلك المعادلة التي اشتقت باستخدام نموذج تحليل الانحدار الخطي للتنبؤ بالعلامة الحقيقية من خـلال التعرف على العلامة الملاحظة وخصائص التوزيع الذي تتبع إليه:

$$t = r_{xx'}.(x)$$

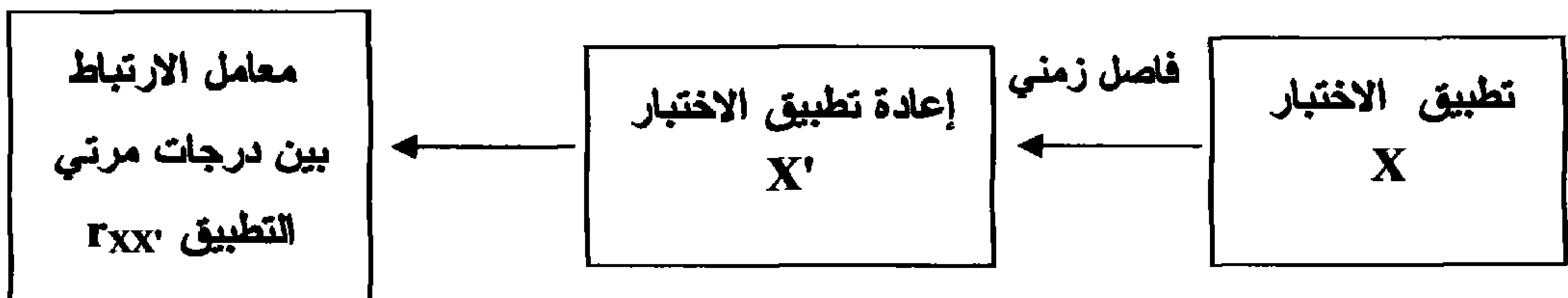
تقدير الثبات Reliability Estimate

بعد أن تم تعريف مفهوم الثبات ومناقشته، يبرز سـؤال حـول مـدى استطاعتنا أن نحصل على تقديرات للثبات المعرّف نظريا. فـإذا أعطينـا مجموعة من الدرجات الملاحظة لمجموعـة مـن الممتحنـين، نـستطيع الحصول على (S_x^2)، أي درجة تباين الدرجات الملاحظة. ومـن خـلال تطبيق المعادلة $r_{xx'} = 1 - \frac{S_e^2}{S_x^2}$، نستطيع الحـصول علـى تقـدير لمعامل الثبات أو تباين الخطأ في الدرجات. ومن البديهي، أن يبدأ الفـرد أولاً بتقدير (r_{xx}) ومن ثم يستخدم المعادلة $S_e = S_x\sqrt{1-r_{xx'}}$ ليقدر قيمة (S_e). وتختلف الطرق المستخدمة لتقدير الثبات في كونها ترجع إلى مصادر مختلفة من الخطأ. ومن الطرق الأكثر شيوعاً هي:

- ثبات الاستقرار **Stability Reliability**
- ثبات التكافؤ **Equivalence Reliability**
- ثبات التكافؤ والاستقرار **Equivalence and Stability**
- مقاييس الاتساق الداخلي **Measures of Internal Consistency**
  - i. التجزئة النصفية **Split-Half**
  - ii. تقديرات كودر- ريتشاردسون **Kuder-Richardson Estimates**
  - iii. معامل ألفا **Coefficient Alpha**
  - iv. طريقة هويت لتحليل التباين **Hoyt's Analysis of Variance**
- ثبات التصحيح **Scoring Reliability**
- ثبات الاختبارات محكية المرجع

ثبات الإستقرار Stability Reliability

يعرف مقياس ثبات الاستقرار في عدد غير قليل من كتب ومراجع القياس النفسي والتربوي بطريقة الاختبار وإعادة الاختبار Test-Retest-Method. ونحصل على معامل ثبات الاختبار بهذه الطريقة،جراء تطبيق الاختبار موضع البحث على مجموعة من الأشخاص، ثم إعادة تطبيق الاختبار ذاته على المجموعة نفسها في وقت لاحق، ويتبع ذلك حساب معامل الارتباط بين درجات أفراد المجموعة على ذلك الاختبار في الفترتين $r_{xx'}$ كما هو مبين في النموذج التالي:



نموذج يوضح إجراءات تقدير معامل ثبات الاستقرار

وجدير بالذكر أنه من النادر ما تستخدم هـذه الطريقـة مـن قبـل المعلمين، أي في حالة الاختبارات التحصيلية، في حين يشيع استـخدامها في حالة الاختبارات النفسية والعقلية وخاصة عند تقييم الاتجاهات والميول والقيم وحتى القدرات والاستعدادات.

في هذا النوع من طرق تقدير الثبات، نستطيع أن نحدد مدى وثوقنـا من إمكانية التعميم من الدرجة التي يحصل عليها الفرد في مرة معينة على الدرجة التي سوف يحصل عليها إذا ما أجري عليه نفس الاختبار في وقت لاحق.

هناك العديد من الفواصل الزمنية (الفترات) الممكنة. ويتنوع تبعا لتلك الفواصل الزمنية تقدير الثبات، وعليه فإن طول هذا الفاصل يجب أن يؤخذ بنظر الاعتبار عند تفسير معامل الثبات. وجدير بالذكر، أن كل تغيير في الدرجة من وقت إلى آخر يعامل معاملة الخطأ (حيث يفتـرض أن تكـون السمة المقاسة ثابتة). ويشبه هذا عملية وزن شخص معين مرتين مختلفتين بنفس المقياس وإرجاع الفرق بين القياسين المسجلين إلى الخطـأ. إذ قـد يمكن أن يرجع هذا الاختلاف إلى طريقة وقوف الشخص على الميزان التي قد تختلف من مرة إلى أخرى، أو ربما يعزى إلى خلل في المقياس (إلى أنه أصبح غير دقيق بين المرتين)، أو قد يعزى ذلك الاختلاف إلى خطأ فـي قراءة أو تسجيل الأرقام، كما يمكن أن يكون الاختلاف ناجماً عـن زيـادة حقيقية في وزن الشخص (عدم ثبات السمة) خلال الوقت. وفي هذا النـوع من التقديرات، لا نستطيع عزل وتحديد مصادر الخطأ التي تسهم في الفرق في الأداء (أو الوزن). حيث أنه ما يقاس فعـلاً هـو الاتـساق فـي أداء الخاضعين للاختبار خلال فترة من الزمن.

كثيراً ما يكون تقدير الثبات شيئاً صعب المنال والتفسير في القيـاس النفسي، لأن العديد من الاختبارات النفسية تفاعلية reactive (Webb and Others, 1981). بمعنى أن أثر القياس نفسه يدفع الشخص على تغيير إجابته

على المتغير المراد قياسه. حيث يحتمل أن تكون تأثيرات الاستجابة من المرة الأولى، على سبيل المثال، مختلفة لدى المستجيبين، مما يؤدي إلى تقليل تقدير الثبات

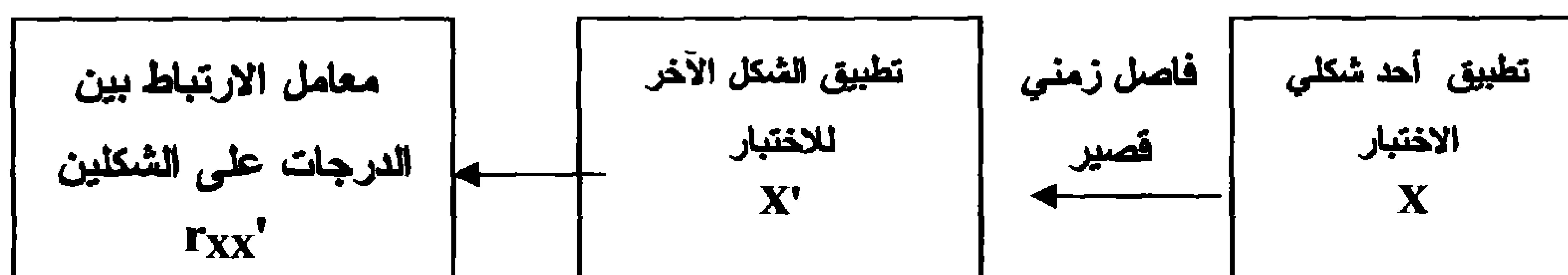
ومن أبرز المشكلات التي تظهر في ثبات الاستقرار ما يدعى بأثر النقل Carry –over effect بين مرتي تطبيق الاختبار. أي أن الأداء على الاختبار في المرة الأولى يؤثر على الأداء في الاختبار عند الإعادة. فإذا كان الفاصل الزمني قصيراً، فسيكون تأثير التذكر memorizing واضحا. بمعنى، أن الطلبة قد يجيبون على السؤال كما فعلوا في المرة السابقة، ليس لكونهم حددوا الحل الصحيح من بين الاختيارات المتوفرة، وإنما لمجرد أنهم تذكروا كيفية تحديدهم للإجابة من المرة السابقة. ويميل تأثير الذاكرة هذا إلى جعل تأثير ثبات الاختبار المعاد عالياً جداً. وغالبا ما تكون مشاكل الذاكرة قليلة الأهمية في الاختبارات في المجال المهاري، ولكنها تكون إشكالية في حالة الاختبارات في المجالين المعرفي والانفعالي. من ناحية أخرى، يظهر أثر النقل بوضوح أكثر بسبب أثر التمرين effect practice وكذلك اختلاف اتجاهات المفحوصين نحو الاختبار في المرة الأخرى. إضافة إلى أن طول الفترة الزمنية الفاصلة بين التطبيقين يكون تأثيرها أكثر ضررا في حالة اختبارات القدرة التي قد تتغير مع تقدم الزمن. والجدول التالي يبين كيفية استخدام معادلة بيرسون لحساب معامل الارتباط الثنائي بين متغيرين كميين (X) و (X'):

	X	X'	X^2	X'^2	XX'
	1	3	1	9	3
	2	6	4	36	12
	4	4	16	16	16
	5	7	25	49	35
المجموع	**12**	**20**	**46**	**110**	**66**

$$r_{xx'}=\frac{n\sum xx'-\sum x\sum x'}{\sqrt{[n\sum x^2-(\sum x)^2][n\sum x'^2-(\sum x')^2]}}=\frac{4(66)-(12)(20)}{\sqrt{[4(46)-(12)^2][4(110)-(20)^2]}}=.6$$

ثبات التكافؤ Reliability Equivalence

على نقيض طريقة الاختبار وإعادة الاختبار، يتم الحصول على تقدير معامل الثبات بطريقة التكافؤ من خلال إعطاء شكلين متكافئين بالمحتوى والمتوسطات والتباينات لاختبار معين لنفس المجموعة وبنفس الوقت (بعد فاصل زمني قصير)، وحساب معامل الارتباط بينهما $r_{xx'}$، كما هو موضح في النموذج التالي:



وبمثل طريقة الاستقرار، فإن هذا الأسلوب لا يستخدم بشكل متكرر من قبل معلمي المدارس أو أعضاء هيئة التدريس في الجامعة. ولا يخفى أن استخدام طريقة التكافؤ يساعد في تحديد مدى الثقة في تعميم درجة شخص ما عندما يحصل عليها إذا خضع لاختبار مشابه ولكن بأسئلة مختلفة.

ولكن إذا اختلف الأداء كثيرا، فإن ذلك خطأ. بمعنى، أن الشخص قد يعرف الإجابة على سؤال في أحد الشكلين ولا يعرف إجابة سؤال مكافئ في الشكل الآخر. بذلك، يعامل الفرق بين الدرجتين على أنه خطأ. وهذا الأسلوب يشبه نوعاً ما عملية وزن شخص ما بمقياسين مختلفين في نفس اليوم. وهنا، لا يحتمل أن نحصل على درجة اختلاف كبيرة (إذا كانت هناك أصلاً أي درجة) ناجمة عن تغيير الوزن، ولكن قد يتواجد الفرق نتيجة للمقياسين المختلفين اللذين استخدما (Mehrens and Lehman,1991).

وجدير بالذكر أنه عند بناء الاختبارات المتكافئة، يجب الانتباه إلـــى ضرورة جعل المقياسين متكافئين من الناحيـــة الإحـــصائية، بمتوســـطات وتباينات وارتباطات متساوية بين الفقرات، إضافة إلـــى ضـــرورة تكـــافؤ محتوى الاختبارين (أحيانا، تعرف الاختبارات المتكافئة بالاختبـــــــارات المتوازية parallel tests).

ومهما يكن من أمر، فإنه يجب اتباع نفـــس جـــدول المواصـــفات المستخدم في بناء الاختبار كما مر في فصول سابقة فـــي كـــلا الـــشكلين. إضافة إلى ما تقدم، يجب أن تكون الفقرات في شكلي الاختبار ذات مستوى صعوبة متشابه وذات نوع متشابه (كالاختيار من متعدد أو فقرات إنـــشائية أو غيرها)، كما أن المفروض أن تكون إرشادات الإجابة في الاختبـــارين متشابهة.

تقدم الأشكال المكافئة للاختبار أكثر من فائدة تعليمية فـــضلا عـــن تقدير الثبات. ففي تقويم المنهاج الدراسي مثلا، يمكن للمدرس أن يجـــري اختباراً بعدياً يغطي المادة التعليمية التي يغطيها الاختبار القبلـــي. عنـــدها يجب استخدام اختبارين متكافئين، أحدهما يستخدم كاختبار قبلـــي والثـــاني اختبار بعدي، بدلاً من إعادة الاختبار نفسه. هذا إضافة لما يفيده اســـتخدام الاختبارات المتكافئة في تقليل أثر التدريس لغرض الاختبار، وكذلك ضبط تأثير الذاكرة التي سبق تناولها في الطريقة السابقة.

ولأن طريقتي تقدير الثبات (الاستقرار والأشكال المتكافئة) تختلفان تماما ويعطي كل منهما نتائج مختلفة، يبرز سؤال حول أي من الطـــريقتين ينصح بالاستخدام؟ وللإجابة على هذا التساؤل، لا بد من تفحـــص الهـــدف جراء تطوير المقياس أو الاختبار. فإذا كان الهدف هو استخدام نتائج الأداء على الاختبار للحصول على توقعات طويلة الأمد، فإننا نستخدم طريقة ثبات الاستقرار. ومثال ذلك، إذا كان موضع اهتمامنا هـــو اختبـــار للاســـتعداد الدراسي في مستوى دراسي ما والذي سيستخدم في التنبؤ بالمعدل التراكمي

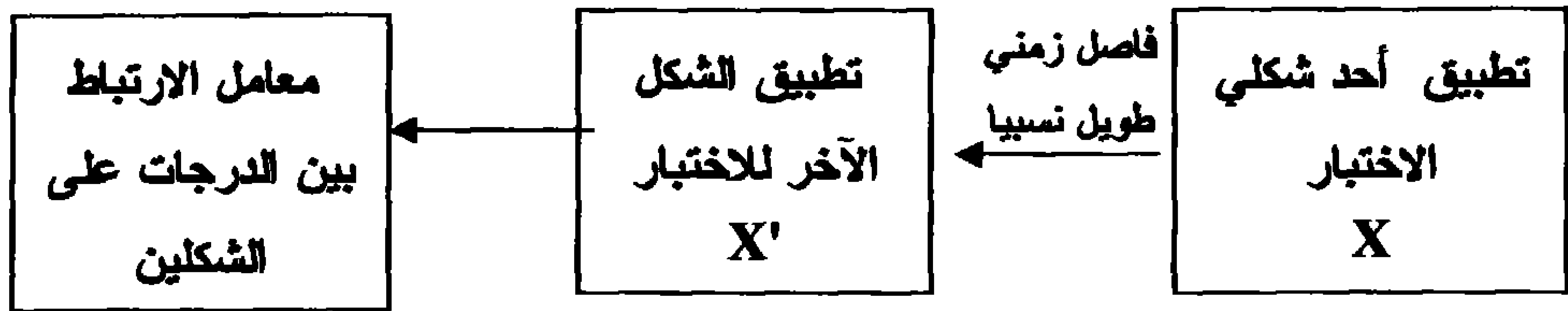
في الدراسة أثناء مرحلة قادمة (الجامعية مثلا). عندها يتوقع أن يكون الاستعداد الدراسي مستقراً مع الزمن وبدرجة كبيرة. وبغير ذلك (إذا كان الاستعداد الدراسي غير ثابت)، فإننا سنكون عاجزين عن الحصول على تنبؤات فعالة وطويلة الأمد.

أما إذا لم تكن الغاية من الاختبار الحصول على تنبؤات بعيدة المدى، وإنما بهدف الوصول إلى استنتاجات حول مدى تمكن الفرد ومعرفته بمحتوى دراسي معين، فسيكون الخيار هو حساب معامل ثبات التكافؤ. وفي هذه الحالة، قلّما نكون معنيين بمدى ثبات المعرفة خلال الزمن، بل سنكون أكثر اهتماما بما يمكن أن نستنتجه أو نعممه من نتائج على العينة على مجال معرفي أوسع.

ثبات التكافؤ والاستقرار Equivalence and Stability Reliability

يهتم بعض الناس أحيانا بتوقعات (تنبؤات) المدى البعيد، وبالاستنتاجات حول مجال معين في المعرفة على حد سواء. وهم في الواقع يعنون بهذين الأمرين أكثر مما يعنون بالثبات. فعلى سبيل المثال، لا يعتمد قياس السمات أو المفاهيم Constructs مثل الذكاء، والإبداع، والعدوانية، أو الولع بالموسيقى على مجموعة متخصصة من الأسئلة. وان كانت كذلك (أي تعتمد على أسئلة كهذه)، لن يكون البناء مفيداً أو مهماً جداً. وكثيراً ما نود أن نعرف ما إذا كانت مجموعة متشابهة من الأسئلة المعطاة في نقاط مختلفة من الوقت، ستعطي نتائج متشابهة.

في هذه الحال، يمكن الحصول على معامل تكافؤ واستقرار من خلال إعطاء شكل من أشكال الاختبار، وبعد فترة زمنية أطول نسبيا، يتم تطبيق الشكل الآخر للاختبار، ويتم حساب معامل الارتباط بين مجموعتي الدرجات كما في النموذج التالي:



نموذج يوضح إجراءات تقدير معامل ثبات الاستقرار – التكافؤ

وتساعد هذه الطريقة في دراسة التغير في الدرجات تبعا لعدم استقرار (ثبات) السمة من ناحية، وكذلك تبعا لضعف تمثيل فقرات الاختبار للمحتوى الدراسي. وبشكل عام، فإن تقدير الثبات هذا يعطي قيما أقل من تلك التي يعطيها أي من الأسلوبين السابقين.

ثبات الاتساق الداخلي Internal Consistency Reliability

تتطلب طرق تقدير الثبات الثلاث التي نوقشت سابقاً بيانات يتم الحصول عليها من جراء جلستين اختباريتين. ويبدو ذلك مشكلا بعض الوقت، عندما لا يكون من الملائم الحصول على هذه البيانات، وخاصة بالنسبة للمعلمين الذين نادراً ما يستخدمون تلك الطرق في تقدير الثبات. وعلى كل حال، من الممكن الحصول على تقديرات ثبات الاتساق الداخلي من خلال تطبيق الاختبار مرة واحدة على مجموعة واحدة من المفحوصين، إضافة إلى أن المعلمين غالبا ما يستخدمون تلك الطرق في اختباراتهم التحصيلية عموما والصفية على وجه الخصوص. وما عدا طريقة التجزئة النصفية، فإن تقديرات الاتساق الداخلي للاختبار هي فعلاً معاملات تجانس لفقرات الاختبار، أو أنها تعكس مدى ترابط الاستجابات على الفقرة الواحدة مع درجة الاختبار الكلية.

ثبات التجزئة النصفية Split-half Reliability

تشبه طريقة التجزئة النصفية لتقدير الثبات نظرياً طريقة ثبات التكافؤ. مع ذلك، تعتبر طريقة التجزئة النصفية مؤشرا لقياس الاتساق الداخلي، لأن الشكلين المتكافئين يؤلفان اختبارا واحدا. وهذان الشكلان بمثابة اختبارين متوازيين. بمعنى أنه، بدلاً عن إجراء شكل بديل للاختبار، نقوم بالتعامل مع اختبار واحد فقط.

وعند تقدير الثبات بطريقة التجزئة النصفية، يتم الحصول على درجة فرعية لكل من النصفين (الفقرات الزوجية مثلا، والفقرات الفردية). ثم يتم حساب معامل الارتباط بين هذين النصفين (r_{hh}). ويكون معامل الارتباط هذا (r_{hh}) تقديراً لثبات اختبار طوله نصف طول الاختبار الأصلي. ولتقدير ما سيكون عليه مستوى ثبات الاختبار كله، نحتاج إلى تطبيق معادلة تنبؤ سبيرمان – براون:

$$r_{xx'} = \frac{2rhh}{1 + rhh}$$

ويشير ($r_{xx'}$) على الثبات المقدر للاختبار ككل، بينما يشير الرمز (r_{hh}) إلى ثبات نصف الاختبار أو هو معامل الارتباط بين نصفي الاختبارين.

تمرين: يبين الجدول التالي درجات عشر مفحوصين على اختبار يتألف من ست فقرات. كذلك علامات هؤلاء الطلبة على الفقرات الفردية والفقرات الزوجية. وتم حساب ثبات التجزئة النصفية للاختبار.

المفحوص	فقرات الاختبار						العلامة على الفقرات		العلامة الكلية
	Q1	Q2	Q3	Q4	Q5	Q6	الفردية	الزوجية	
1	0	0	0	0	0	0	0	0	0
2	0	0	0	0	1	0	1	0	1
3	1	0	1	1	1	0	3	1	4
4	1	1	1	1	1	1	3	3	6
5	1	1	1	1	1	1	3	3	6
6	0	0	1	0	0	0	1	0	1
7	0	0	1	1	1	0	2	1	3
8	0	0	0	1	0	0	0	1	1
9	1	0	1	1	1	0	3	1	4
10	0	1	0	1	0	1	0	3	3

تم حساب معامل الارتباط بين العلامات على الفقرات الفردية والعلامات على الفقرات الزوجية وكان (0.34). أي أن $r_{hh} = 0.34$. وعليه، يكون معامل ثبات التجزئة النصفية يساوي (0.51) كما في المعادلة التالية:

$$r_{xx'} = \frac{2rhh}{1 + rhh} = \frac{2(0.34)}{1 + 0.34} = 0.51$$

مثال: إذا كان معامل الارتباط بين نصفي اختبار ما (0.60) فما مقدار معامل ثبات الاختبار كاملا ؟

$$r_{xx'} = \frac{2(0.60)}{1+0.60} = \frac{1.20}{1.60} = 0.75$$

وتكمن فائدة طريقة القسمة إلى نصفين في أننا نحتاج إلى تطبيقه على التلاميذ لمرة واحدة فقط بالرغم من أنها تفترض أن تباين نصفي الاختبار متساوية. إذ أنه إن لم تكن التباينات متساوية، فإن الثبات المقدر للاختبار ككل سيكون أكبر من ذلك الذي حصلنا عليه من الطرق الأخرى للإتساق

الداخلي. وعليه، فإن أحد المشكلات الموجودة في هذه الطريقة هي كيفية قسمة الاختبار إلى قسمين، ويمكن التعامل مع هذه المشكلة بأساليب متنوعة.

لكن إذا حاول الفرد حقاً أن يجعل النصفين متكافئين (ومتوازيين)، فإن هذا يتطلب منه كل الجهود الكفيلة ببناء شكلين متكافئين (ماعدا أن عدد فقراتها يساوي نصف الفقرات المطلوبة). ويتم تقسيم الاختبار إلى جزأين وفقاً لخطة مسبقة من أجل جعل النصفين متساويين. وفي غالب الأحيان لا يكون تباينا نصفي الاختبار متساويين أو من نوع Tau-Equivalent. الأمر الذي أدى استبدال معادلة سبيرمان–براون بأحد المعادلتين التاليتين هما:

1. معادلة جوتمان (Guttman) كما في الصيغة التالية:

$$r_{xx'} = 2\{1 - (\frac{S_1^{\ 2} + S_2^{\ 2}}{S_x^{\ 2}})\}$$

حيث تشير $(S_1)^2$ و $(S_2)^2$ إلى تباين النصف الأول للاختبار وتباين النصف الثاني له على الترتيب. بينما يرمز $(S_x)^2$ إلى تباين الدرجات على الاختبار ككل. من ناحية أخرى، إذا تم تقسيم الاختبار إلى أكثر من قسمين متكافئين (ثلاثة فأكثر)، فإنه يمكن كتابة معادلة جوتمان على النحو التالي:

$$r_{xx'} = 2\{1 - (\frac{S_1^{\ 2} + S_2^{\ 2} + S_3^2}{S_x^{\ 2}})\}$$

2. معادلة رلون Rulon's Equation

ويتم هنا استخدام الفرق بين الدرجات على النصفين كما في الصيغة التالية:

$$r_{xx'} = 1 - \frac{S_D^{\ 2}}{S_x^{\ 2}}$$

حيث يشير الرمز S_D^2 إلى تباين الفروق بين في العلامات على نصفي الاختبار.

تمرين: يبين الجدول التالي توزيعا لدرجات عشر طلاب على اختبار يتألف من 40 فقرة انتقائية، والمطلوب حساب معامل ثبات الاختبار باستخدام كل من معادلة سبيرمان –براون ومعادلة جوتمان ومعادلة رلون من ثم المقارنة بينها:

رقم المفحوص	درجة النصف الأول	درجة النصف الثاني	درجة الاختبار ككل	D
01	19	20	39	-1
02	17	15	32	2
03	11	13	34	-2
04	9	zero	9	9
05	8	7	15	1
06	19	17	36	2
07	13	11	24	2
08	9	9	18	0
09	15	18	33	-3
10	10	11	21	-1
S^2	18	35	102.3	11.21

تم حساب معامل معامل الارتباط بين نصفي الاختبار وكـــان يـــساوي (0.89). وكان ثبات التجزئة النصفية

باستخدام معادلة سبيرمان –براون:
$$r_{xx'} = \frac{2(0.89)}{1+0.89} = \frac{1.78}{1.89} = 0.94$$

وباستخدام معادلة جوتمان:
$$r_{xx'} = 2\{1 - (\frac{18 + 35}{102.3})\} = 0.96$$

وباستخدام معادلة رلون:
$$r_{xx'} = 1 - \frac{11.21}{102.3} = 0.89$$

ويعزى الفرق بين قيم المعاملات بسبب اختلاف تباين نصفي الاختبار (18) و (35) كما ظهر في الجدول. ومما لا شك فيه أن القيم تقترب من بعــض أكثر، كلما قل الفرق بين تباين نصفي الاختبار كما في المثال التالي:

تمرين: يبين الجدول التالي توزيعا لدرجات عشر طلاب على اختبار يتألف من 20 فقرة انتقائية، والمطلوب حساب معامل ثبات الاختبار باستخدام كل

من معادلة سبيرمان –براون ومعادلة جوتمان ورلون ومن ثـم المقارنـة بينها:

رقم المفحوص	درجة النصف الأول	درجة النصف الثاني	درجة الاختبار ككل	D
01	10	10	20	0
02	8	9	17	1-
03	5	4	9	1
04	4	5	9	-1
05	4	7	11	-3
06	9	8	17	1
07	7	4	11	3
08	4	4	8	0
09	8	8	16	0
10	3	3	6	0
S^2	6.2	6.2	22.3	2.44

تم حساب معامل الارتباط بين نصفي الاختبار وكان يساوي (0.8). وكان ثبات التجزئة النصفية باستخدام:

معادلة سبيرمان-براون: $r_{xx'} = \frac{2(0.8)}{1+0.8} = \frac{1.6}{1.8} = 0.88$

وباستخدام معادلة جوتمان: $r_{xx'} = 2\{1 - (\frac{6.2 + 6.2}{22.3})\} = 0.88$

وباستخدام معادلة رلون: $r_{xx'} = 1 - \frac{2.44}{22.3} = 0.89$

من هنا يلاحظ أن معادلات سبيرمان-براون، وجوتمان، ورلون تعطي نفس قيمة معامل الثبات عندما يتساوى تباين نصفي الاختبار. من ناحية ثانيـة، فإن مقدار معامل ثبات الاختبار يزداد كلما يزداد معامـل الارتبـاط بـين نصفيه. فإذا تلاشى معامل الارتباط بين نصفيه (r_{hh}=0.0)، فإن معامل ثبات

الاختبار ككل يساوي صفرا. وإذا كانت العلاقة بينهما تامة، يصل معامل ثبات الاختبار درجة الكمال ($r_{xx'}=1$) وهكذا.

قديرات كودر–ريتشاردسون Kuder- Richardson Estimates

إذا تم تصحيح الفقرات بشكل ثنائي (إما صحيحة =1 ، أو خاطئة =0)، فإن إحدى طرق تجنب مشاكل كيفية قسمة الاختبار هي استخدام معادلات كودر– ريتشاردسون. ويمكن اعتبار هذه المعادلات ممثلاً لمعدل معامل الارتباط الحاصل من جميع التقديرات الممكنة للثبات المقسمة إلى نصفين. ومن المعادلات شائعة الاستخدام في تلك التقديرات معادلتا كودر– ريتشاردسون 20، 21 (KR-20) و (KR-21). والمعادلة التالية هي معادلة كودر–ريتشاردسون 20 لحساب الثبات:

$$r_{xx'} = \frac{n}{n-1}\left(1 - \frac{\sum pq}{S_x^2}\right)$$

أما معادلة كودر– ريتشاردسون 21 ، فيمكن عرضها كما في المعادلة التالية:

$$r_{xx'} = \frac{n}{n-1}\left(1 - \frac{\overline{X}(n-\overline{X})}{nS_x^2}\right)$$

وتشير n إلى عدد الفقرات في الاختيار. بينما تشير p إلى نسبة الأفراد الذين أجابوا بشكل صحيح على الفقرة (صعوبة الفقرة). بينما تشير q إلى نسبة الأفراد الذين أجابوا على الفقرة بشكل خاطئ ($q= 1-p$)، فإذا كان: ($p=0.20$)، فإن ($q=0.80$). وتبين pq تباين الفقرة المصححة بشكل ثنائي (صح أو خطأ). ويدل S_x^2 على تباين مجمل الاختبار. والرمز $\overline{X}$ هو متوسط مجمل العلامات الكلية على الاختبار.

والفرق بين معادلة **KR-20** و **KR-21** هو أن الأخيرة تفترض أن جميع الفقرات ذات مستوى صعوبة واحد، بمعنى أن قيمة (p) ثابتة لكل الفقرات. وإن لم يتحقق هذا الافتراض فإن **(KR-21)** ستعطي تقديراً للثبات أقل

بقليل. وبالرغم من أن كلا المعادلتين كثير الاستخدام لدى ناشري الاختبارات، كما يبدو أن هذه المعادلة (**KR-21**) تبدو مغرية للاستخدام من قبل المهتمين لكونها تتطلب عمليات حسابية أقل مما تتطلبه (KR-20). وإذا توفر لدينا عدد الفقرات (n) في الاختبار، فإننا نحتاج فقط إلى حساب متوسط وتباين الدرجات الكلية على الاختبار، ونعوض هذه القيم الثلاثة في المعادلة، ونجد معامل الثبات.

وبالرغم من أن كلا المعادلتين يعطي حدوداً للثبات، فإن التقدير العالي KR-20) هو الأفضل دائماً ويجب استخدامه كلما كان ذلك ممكناً. وعلى الرغم من أن جهدا كبيرا يبذله مطور الاختبار في حساب معاملات صعوبة الفقرات، إلا أن هذه المعلومات مفيدة جداً، الأمر الذي يجعلنا نتمنى أن يقوم جميع المدرسين بحسابها على أية حال

.

ثبات كرونباخ ألفا Cronbach Alpha Reliability (α)

هي الطريقة التي اقترحها وطورها كرونباخ (Cronbach) عام (1951) لتقدير ثبات الاتساق الداخلي للاختبار، وهي تعميم لمعادلة (KR-20) عندما لا يتم تصحيح الفقرات بشكل ثنائي. ويشيع استخدام هذه الطريقة في تقدير ثبات مقاييس الاتجاهات واستطلاع الرأي وفي مقاييس الشخصية، وفي حالة الاختبارات التحصيلية الصياغية. وجدير بالذكر أن طريقة ألفا تعطي الحد الأدنى للقيمة التقديرية لمعامل ثبات درجات الاختبارات. فإذا كانت قيمة ألفا مرتفعة، فهذا يدل بالفعل على ثبات الاختبار. أما إذا كانت منخفضة فربما يدل على أن الثبات يمكن أن تكون قيمته أكبر من ذلك باستخدام الطرق الأخرى.

من ناحية رياضية، تظهر معادلة ألفا على أنها صيغة أخرى لمعادلة (KR-20)، سوى أن مجموع التباين ($\sum pq$) قد استبدل بمجموع تباينات فقرات الاختبار $\sum S_i^2$ كما في المعادلة التالية:

$$\alpha = \frac{n}{n-1}\left(1 - \frac{\sum S_i^2}{S_x^2}\right)$$

حيث أن S_i^2 هو تباين الفقرة الواحدة. أما الرموز الأخرى فقد تم توضيحها سابقاً. وجدير بالذكر أن طريقة ألفا تتمتع بأهمية خاصة، كونها تستخدم في حساب معامل ثبات الاختبارات المقالية والموضوعية على حد سواء. إذ يمكن أن تصحح درجة المفحوص على كل سؤال باستخدام مدى من القيم.

من ناحية أخرى، فإن قيمة معامل ألفا تساوي متوسط القيم التقديرية لمعامل ثبات كل من نصفي الاختبار لجميع طرق التجزئة النصفية الممكنة، كما هو الحال في معامل التجانس لكودر-ريتشاردسون.

تمرين: يعرض الجدول التالي استجابات 15 طالبا على اختبار يتألف من 10 فقرات انتقائية. كما ظهر في العمود الأخير الدرجات الكلية لكل مفحوص. وسيتم حساب معامل ثبات الاتساق الداخلي للاختبار باستخدام معادلة كرونباخ ألفا ومعادلتي كودر-ريتشاردسون (20) و (21):

رقم المفحوص	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	الدرجة الكلية
01	1	1	1	1	1	1	1	0	1	1	9
02	1	1	1	1	1	1	1	0	1	1	9
03	1	1	1	1	1	1	1	0	1	0	8
04	1	0	0	1	1	1	1	0	1	1	7
05	1	1	1	0	1	1	0	1	0	0	6
06	0	0	0	1	1	1	0	1	0	1	5
07	0	1	1	1	0	0	0	1	1	0	5
08	1	1	1	1	0	0	0	0	0	0	4
09	0	0	0	0	1	0	0	1	1	1	4
10	1	0	0	1	1	0	0	0	0	0	3
11	1	1	0	0	0	0	0	0	0	1	3
12	0	1	0	0	0	0	0	1	0	0	2
13	0	0	0	0	0	0	1	0	0	0	1
14	0	1	0	0	0	0	0	0	0	0	1
P	.57	.64	.42	.57	.57	.42	.35	.35	.42	.42	
q	.43	.36	.58	.43	.43	.58	.65	.65	.58	.58	
pq	.245	.23	.24	.245	.245	.24	.23	.23	.24	.24	
S^2	.26	.24	.264	.26	.26	.264	.247	.247	.264	.264	7.4

وقد تم حساب المتوسط الحسابي للعلامات الكلية على الاختبار وكان(4.8) وتباين الدرجات الكلية وكان (7.4). وكان مجموع حاصل ضرب المعامل (p) بالمعامل (q) يساوي (2.4).

معامل ثبات كرونباخ ألفا:

$$\alpha = \frac{10}{10-1}\left(1-\frac{2.6}{7.4}\right) = 0.72$$

معامل (KR-20):

$$r_{xx'} = \frac{10}{10-1}\left(1-\frac{2.4}{7.4}\right) = 0.75$$

معامل (KR-21):

$$r_{xx'} = \frac{10}{10-1}\left(1-\frac{4.8\,(10-4.8)}{10\,(7.4)}\right) = 0.73$$

يبدو أن معاملات الثبات المحسوبة باستخدام معادلة ألفا و معادلتي كودر-ريتشاردسون متقاربة بدرجة كبيرة بسبب تقارب معاملات صعوبة فقرات الاختبار كما هو مبين. هذا على الرغم من أن معامل ثبات كرونباخ ألفا يشكل الحد الأدنى لقيم معاملات ثبات الاتساق الداخلي.

مثال: إذا كان لدينا مقياسا للاتجاهات يتضمن 6 عبارات على مقياس ثلاثي التدريج (موافق=3)، (محايد= 2)، (غير موافق=1)، وتم تطبيقه على خمسة أشخاص. ما مقدار ثبات ألفا لهذا المقياس إذا كانت درجاتهم كما في الجدول التالي:

الشخص	Q1	Q2	Q3	Q4	Q5	Q6	المجموع
1	2	1	3	2	3	1	12
2	3	2	3	3	2	3	16
3	2	1	2	3	1	2	11
4	2	1	1	1	1	1	7
5	3	2	3	2	3	2	15
S^2	.3	.3	.8	.7	1	.7	12.7

وقد تم حساب القيمة المتوسطة للدرجات على المقياس وكانت (12.2)، وكان تباين الدرجات الكلية (3.8). ثم نطبق معادلة ألفا على النحو التالي:

$$\alpha = \frac{6}{6-1}\left(1 - \frac{3.8}{12.7}\right) = 0.84$$

أسلوب هويت (Hoyt) في تحليل التباين

تتطابق النتائج التي يعطيها أسلوب هويت في تحليل التباين مع تلك التي تعطيها معادلة (KR-20) عندما يتم التعامل مع الفقرات التي تصحح ثنائيا، أو طريقة ألفا. والمعادلة التالية تعطي معامل ثبات هويت باستخدام تحليل التباين الثنائي. على اعتبار أن الأشخاص examinees

والفقرات items هما المتغيران التصنيفيان، بينما تكون العلامات على الفقرات هي المتغير التابع:

$$r_{xx'} = \frac{MS_{examinees} - MS_{interaction}}{MS_{examinees}}$$

ففي التمرين المتعلق بإيجاد معامل ثبات التجزئة النصفية، كان لدينا علامات 10 من الطلبة على ست فقرات. وقد أمكن حساب معامل ثبات تلك الفقرات بتطبيق المعادلة بعد استخراج مكوناتها جراء تحليل التباين الثنائي: **Two-Way-ANOVA (6x10)**.

والجدير بالذكر أنه يمكن استخدام برنامج SPSS وذلك بإدخال البيانات على هيئة ثلاثة أعمدة: واحد يمثل أرقام المفحوصين Examinees ، وعمود يمثل الفقرات Item، وعمود يحوي العلامات على الفقرات Score. بعد ذلك يصار إلى إدخال score في خانة المتغير التابع dependent variable ويتم تحريك المتغيرين Examinees و Item إلى خانة المتغيرات التصنيفية Fixed effect. ومن ثم نحصل على النتيجة التالية:

ANOVA[a,b]

			Unique Method		
			Sum of Squares	df	Mean Square
score	Main Effects	(Combined)	8.300	14	.593
		examinee	6.817	9	.757
		item	1.483	5	.297
	2-Way Interactions	examinee * item	6.683	45	.149
	Model		14.983	59	.254
	Residual		.000	0	.
	Total		14.983	59	.254

a. score by examinee, item

b. All effects entered simultaneously

ومن مطالعة هذا الجدول نلاحظ أن ($MS_{examinee}$=.757)، و(MS_{item}= .297)، و ($MS_{interaction}$=.149). وعليه، يكون معامل الثبات يساوي (0.8039) هكذا:

$$r_{xx'} = \frac{MS_{examinees} - MS_{interaction}}{MS_{examinees}} = \frac{.757 - .149}{.757} = 0.803$$

ثبات الفرق بين درجتي اختبارين Reliability of the differences

قد يتم أحيانا مقارنة درجة شخص في اختبارين أو أكثر، كمقارنة درجته على مقياس تقدير الذات مع درجته على مقياس التوافق الشخصي والاجتماعي، أو ربما نلجأ إلى مقارنة درجة طالب على مادة ما بدرجته على مادة أخرى، أو قد يصار إلى مقارنة أداء متدرب قبل البرنامج وأدائه بعد البرنامج، عند ذلك لا بد من دراسة ثبات الفرق بين الدرجتين. ويتم هذا كما في المعادلة التالية:

$$r_{xx'} = \frac{\frac{r_{11} + r_{22}}{2} - r_{12}}{1 - r_{12}}$$

حيث يشير r_{11} و r_{22} إلى معامل ثبات الاختبار الأول ومعامل ثبات الاختبار الثاني على الترتيب. بينما يشير الرمز r_{12} إلى معامل الارتباط بين الاختبارين.

مثال: طبق اختباران على مجموعة من الطلبة، ثبات الاختبار الأول (.8) وثبات الاختبار الثاني (.9) ومعامل الارتباط بين الاختبارين (.40) فما معامل ثبات الفرق بين درجتي الاختبارين؟

$$r_{xx'} = \frac{\frac{.80 + .90}{2} - .40}{1 - .40} = 0.75$$

ويلاحظ أن القيمة الناتجة أقل من قيمة معامل ثبات درجات أي من الاختبارين. وكلما زاد معامل ثبات كل اختبار وقل الارتباط بينهما، يزداد معامل ثبات الفرق بين الدرجتين. وعندما يكون معامل ثبات كل اختبار يساوي (1) فإن ثبات الفرق بين الدرجتين يساوي (1) أيضا مهما بلغت قيمة معامل الارتباط بين الاختبارين. ويمكن تطبيق ذلك على المعادلة السابقة كما هو مبين تاليا:

$$r_{xx'} = \frac{\frac{1+1}{2} - r_{12}}{1 - r_{12}} = \frac{1 - r_{12}}{1 - r_{12}} = 1$$

ثبات التصحيح Scoring Reliability

عرفنا بداية أن هناك مصادر مختلفة تؤدي إلى وجود أخطاء عشوائية في درجات الممتحنين. كما عرفنا أيضا أن بعضاً من طرق تقدير الثبات ربما يسمح لمصدر أو أكثر من مصادر الخطأ بالحدوث. فمثلا ، يؤدي عدم استقرار السمة إلى ظهور الخطأ العشوائي إذا تم استخدام طريقة الاستقرار في تقدير ثباتها. ويمكن أن يقع خطأ المعاينة إذا ما استخدمنا طرق ثبات الاتساق الداخلي. كذلك يمكن أن يقع خطأ القائم بالاختبار إذا تم تطبيق الاختبار نفسه مرتين، أو تطبيق اختبارين متكافئين معا.

أما عند تصحيح الأداء على الاختبارات المقالية، أو تحليل السلوك على الأدوات الإسقاطية للشخصية، وتقدير الدرجات على مقاييس التصنيف، هناك احتمالية حدوث خطأ في الدرجة يرجع إلى الأشخاص الذين قاموا بالتصحيح أو التصنيف.

قد يتم تكليف مصحح واحد بتصحيح أوراق الامتحان مرتين ومن ثم يتم تقدير درجة الارتباط بين الدرجات في المرتين. ويعرف هذا بثبات المصحح Intra-rater Reliability. أما عندما يطلب من مصححين اثنين

ليقوما بالعملية، بحيث يصحح كل واحد الامتحان ثم يصححه زميله الآخر بشكل مستقل. ومن ثم يتم حساب درجة الارتباط بين مجموعتي الدرجات. وهذا يعرف بثبات المصححين Inter-rater Reliability.

فإذا أعطيت عينة من الأوراق للتصحيح بشكل مستقل من قبل شخصين مختلفين، يمكن استخدام معامل ارتباط بيرسون (r) لغرض تقدير ثبات الدرجة التي يعطيها كل قارئ على حدة. وإذا أردنا معرفة ثبات المجموع (أو المعدل) لدرجتي الشخصين المصححين، فعلى المعلم أن يستخدم معادلة تنبؤ سبيرمان – براون. في هذه الحالة سيكون (r_{hh}) معامل الارتباط بين مجموعتي الدرجات، ويكون ($r_{xx'}$) هو معامل الثبات الذي تم تقديره لدرجات المجموع (أو المعدل).

وفي بعض الأحيان، فإن مصحح واحد أو مصححين (محكمين) اثنين قد لا يوفران بيانات ثابتة بما فيه الكفاية. عندها، إذا أستخدم أكثر من مصححين أو محكمين.

مقارنة بين طرق تقدير الثبات

تشير البيانات الواردة في الجدول التالي إلى استعراض طرق تقدير الثبات ومصادر الخطأ المحتملة في كل منها. ويلاحظ أن مصادر الخطأ تتكرر بشكل واضح في حالة تقدير الثبات باستخدام طريقتي التكافؤ والاستقرار أكثر مما هو في أي طريقة أخرى. الأمر الذي يؤدي في النهاية إلى أن معامل الثبات المقدر بهاتين الطريقتين أكثر نزوعاً لأن يكون أقل منه في الحالات الأخرى. وهذا يشير إلى ضرورة النظر إلى طرق تقدير الثبات التي تستخدم عند تطوير أو تقنين المقاييس، وليس الاكتفاء فقط بالقيمة الرقمية لمعامل الثبات.

مصادر الخطأ المحتمل حدوثها في طرق تقدير الثبات

المصادر المحتملة للخطأ	طريقة تقدير الثبات				
	الاستقرار	التكافؤ	الاستقرار والتكافؤ	الاتساق الداخلي	ثبات التصحيح
عدم استقرار السمة	X		X		
خطأ معاينة		X	X	X	
إدارة الاختبار	X	X	X		
خطأ عشوائي	X	X	X	X	X
خطأ التصحيح					X

يتضح من قراءة هذا الجدول، أن الخطأ العشوائي يمكن أن يكون ذا تأثير على كافة طرق تقدير الثبات.

العوامل التي تؤثر في مقدار معامل الثبات

يتأثر معامل ثبات أداة القياس بعدد من العوامل، منها ما يتعلق بأداة القياس نفسها ، وبعضها يتعلق بخصائص الأفراد الذين يطبق عليهم الأداة موضع البحث. وفي هذا الفصل، سيتم مناقشة تأثير عدد من تلك العوامل، مثل طول الأداة أو عدد فقراتها، طبيعة الأداة (اختبار سرعة، اختبار قدرة)، مدى تجانس أفراد المجموعة، مدى صعوبة الفقرات، إضافة إلى الموضوعية.

طول الأداة

بشكل عام، تعطي الاختبارات الأكثر طولاً درجات أكثر ثباتاً. ويصح هذا بدرجة كبيرة لأن الأخطاء العشوائية الموجبة والسالبة في الاختبار ستحظى بفرصة أن يلغي أحدهما الآخر (Mehrens and Lehman, 1991)، بذلك تقترب الدرجة الملاحظة (X) من الدرجة الحقيقية (T). وبنفس الطريقة، تعطي الاختبارات القصيرة أو الفرعية درجات أقل ثباتاً.

وهذا بطبيعة الحال، ينطبق على كافة أدوات القياس وليس فقط على اختبارات التحصيل.

من هنا، كان المفروض أن يدرك الفرد تلك الحقيقة عندما يعد الاختبارات من ناحية، أو عندما يفسر درجات المفحوصين على تلك الاختبارات من ناحية أخرى. إذ يقدر مستوى الثبات الذي يمكن أن يتمتع به الاختبار في حالة كونه قصيرا أو غير ذلك باستخدام معادلة تنبؤ سبيرمان – براون. وعندما ناقشنا طريقة القسمة إلى نصفين لتقدير الثبات، قمنا بتوضيح حالة خاصة من حالات معادلة تنبؤ سبيرمان – براون. أما التعبير العام عن هذه المعادلة فهو:

$$r_{xx'} = \frac{kr}{1+(k-1)r}$$

حيث يشير ($r_{xx'}$) إلى معامل الثبات المتوقع لاختبار طوله عدد مرات الاختبار الأصلي. ويدل الحرف (r) على ثبات الاختبار الأصلي، والحرف K يشير إلى نسبة عدد الفقرات في الاختبار الجديد إلى عدد الفقرات في الاختبار الأصلي.

مثال: إذا كان للاختبار ثبات مقداره (0.60). ثم جعلنا طول الاختبار ثلاثة أمثال طوله الأصلي (k=3) (مثل الذهاب من اختبار بعشرين فقرة إلى أن يصبح طوله ستين فقرة)، فإننا سنتوقع أن يكون ثبات الاختبار المطول كما يلي:

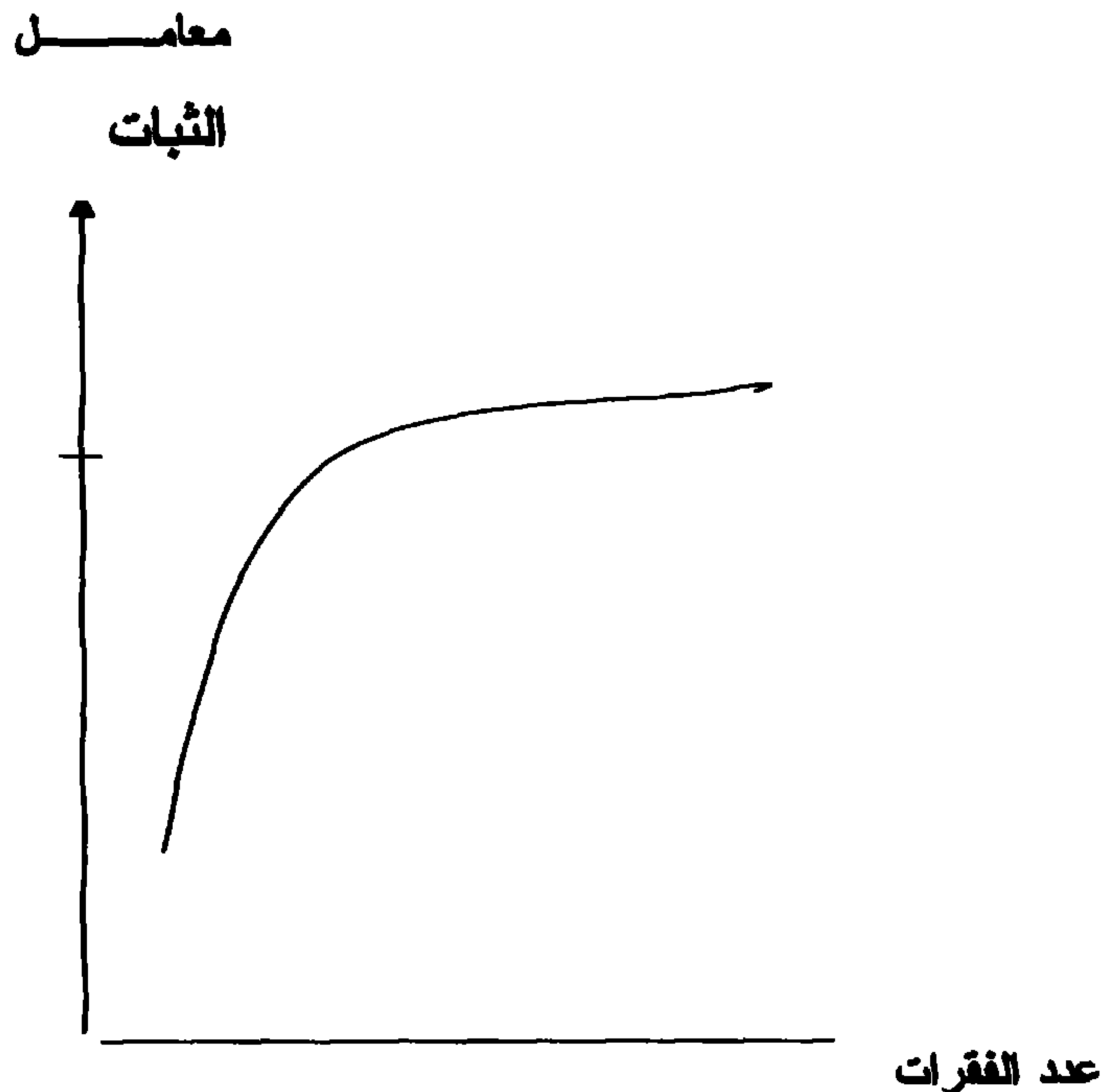
$$r_{xx'} = \frac{3(0.60)}{1+(3-1)(0.60)} = \frac{1.8}{2.2} \cong .82$$

وكما ذكرنا سابقاً، عندما كانت قيمة (K) تساوي (2) (كما في طريقة القسمة على نصفين لتقدير الثبات) فإن معادلة سبيرمان – براون ستفترض أن الاختبارين الفرعيين متوازيان. ونستطيع أن نعبر عن هذا الافتراض بطريقة أخرى، هي أن الفقرات المضافة إلى الاختبار يجب أن

تكون مكافئة للفقرات الموجودة أصلاً فيه، وأن تلك الفقـــرات المـــضافة لا تسبب أية عوامل أخرى كالتعب والإعياء.

ومثلما تجعل عملية إضافة فقرات مكافئة أو موازية إلى الاختبـــار، تجعل درجة الاختبار أكثر ثباتاً، فإن عملية حذف الفقرات المكافئة تجعـــل الاختبار أقل ثباتاً.

من ناحية أخرى، قد يتمتع اختبار ما بدرجة ثبـــات عاليـــة، إلا أنـــه أطول مما نستطيع أن نستخدم. ويمكن استخدام المعادلة في تقـــدير ذلـــك. فمثلا، إذا أردنا أن نعرف مقدار معامل ثبات اختبار عندما يـــصبح طولـــه نصف طوله الأصلي، نستخدم المعادلة نفسها وتكون(K=0.50). والـــشكل التالي يبين علاقة مقدار قيمة معامل الثبات بدلالة عدد فقراته:



ويمكن حساب قيمة k عن طريق استخدام المعادلة التالية:

$$k = \frac{r_{xx'}(1 - r)}{r(1 - r_{xx'})}$$

مثال: إذا كان لديك اختبار تحصيلي في الرياضيات يتألف من 30 فقــرة، ومعامل ثبات مقداره (0.60). وبعد عرضه على محكمين محترفين، أشاروا بضرورة تطويره ليصبح معامل ثباته أفضل وبما لا يقل عن (0.70)، فكم فقرة يجب إضافتها للإختبار لتحقيق هذا المستوى من الثبات؟

الحل: يتم تطبيق المعادلة لحساب مقدار k على النحو التالي:

$$k = \frac{r_{xx'}(1 - r)}{r(1 - r_{xx'})} = \frac{0.70\ (1 - 0.60)}{0.60\ (1 - 0.70)} = 1.5$$

بناء على ذلك، يتم تكبير الاختبار مرة ونصف. أي يصبح طول الاختبــار 45 فقرة ($30 \times 1.5 = 45$). وعليه يجب إضافة (15) فقرة جديدة للاختبار.

تجانس مجموعة المفحوصين

يزداد معامل ثبات الاختبار إذا ازدادت درجــة تبــاين مجموعــة المفحوصين، شريطة ضبط العوامل الأخرى المؤثرة في ثبات الاختبــار. بمعنى أنه كلما كان أداء أفراد مجموعة المفحوصين متباينا أكثر، كلما زاد مقدار معامل ثبات الاختبار. ويمكن إدراك تلك العلاقة من خلال تفحــص المعادلة:

$$r_{xx'} = 1 - \frac{S_e^2}{S_x^2}$$

وذلك عندما يزداد تباين درجات المفحوصين S_x فإن معامل الثبات سيزداد.

وعندما يتم تطوير اختبار ما على مجموعة من المتفوقين فقط أو ضعيفي التحصيل فقط، فإن معامل الثبات سيكون منخفضا. الأمر، الذي يشير إلى وجوب أن يتم تطوير أداة القياس بعد تطبيقها على عينة ممثلة للمجتمع المنوي تعميم نتائج الاختبار عليه. وذلك بأن تغطي العينة كافة طبقات المجتمع ومستوياته.

صعوبة الفقرات

بما أن معامل الثبات يعتمد على درجة التباين في درجات المفحوصين (كما في المعادلة السابقة)، فإن معامل الثبات يزداد تبعا للعوامل التي تحسن مستوى ذلك التباين. وهذا يتم من خلال الأداء على الفقرات التي تتمتع بمعاملات صعوبة متوسطة. إذ أن الفقرة السهلة جدا أو الصعبة جدا، لا تبرز الفروق الفردية، وهي لا تظهر التباين بين درجات المفحوصين. الأمر الذي يؤدي إلى جعل معامل الثبات منخفضا.

الموضوعية

بعد ضبط العوامل الأخرى المؤثرة في معامل الثبات، فإن الفقرات التي تصحح بموضوعية عالية تؤدي إلى تحسين (زيادة) تقدير معامل الثبات. وهذا يشير إلى أن ثبات الاختبار الذي يتألف من فقرات انتقائية أعلى ثباتا عموما من الاختبار الذي يتألف من فقرات انشائية شريطة تساوي العوامل المؤثرة الأخرى.

ثبات الاختبارت محكية المرجع

تعتمد طرق تقدير ثبات الاختبارات محكية المرجع على كيفية استخدام الاختبار، بعكس طرق تقدير ثبات الاختبارات معيارية المرجع التي تعتمد على مفهوم معامل الارتباط بين متغيرين. إذ يتم استخدام فكرة معامل الارتباط في تقدير ثبات الاستقرار أو التكافؤ أو حتى ثبات الاتساق

الداخلي للاختبارات معيارية المرجع. وعليه هناك مجموعتان من طرق تقدير ثبات الاختبارات محكية المرجع. الأولى تتعلق بتقدير درجات المفحوصين في نطاق سلوكي معين، والمهم هنا تقليل الخطأ الناجم عن الفروق التقديرية لدرجات الأفراد في النطاق السلوكي ودرجات النطاق إلى أقل تقدير، ومن هذه الطرق معامل ثبات ليفنجستون Livingston Index. والثانية تتعلق بتصنيف الأفراد في مجموعات حسب تمكنهم من نطاق سلوكي معين، والمهم هنا تحديد الأخطاء الناجمة عن عدم اتساق التصنيف إذا طبق اختباران متوازيان.

ومهما يكن من أمر، فعند الحديث عن ثبات الاختبارات محكية المرجع لا بد من معرفة مفهومي الاختبارات المتوازية Parallel Tests ودرجات القطع Cut-off Scores. والإختبارات المتوازية هي مجموعة من الأدوات التي تشتق فقراتها من نطاق سلوكي واحد ويتمتع كل منها بمؤشرات متشابهة لكل من الأهداف والمتوسطات الحسابية والتباينات والتوزيع البياني، وعندها تعرف تلك الاختبارات بالاختبارات المتوازية كلاسيكيا Classically Parallel Tests. بينما تعرف الاختبارات المتوازية التي تكون فقراتها مشتقة عشوائيا من النطاق الشامل للفقرات ولا تتطلب تحقق أي من الفروض السابقة، وهذه تعرف بالاختبارات المتوازية عشوائيا Randomly Parallel Tests.

أما درجات القطع، فيجب أن يبرر تقديرها في كل اختبار وهي تختلف طبعا باختلاف درجة الدقة في قرار التصنيف للأفراد موضع البحث. ولا شك أن عوامل عدة تدخل في تحديد تلك الدرجات. ومهما يكن من أمر، فالعبرة ليست باتساق القرار بل بدقته.

معامل ليفنجستون Livingston Index

اعتمد ليفنجستون في اشتقاقه لهذا المؤشر على أسس النظرية الكلاسيكية في تقدير ثبات الاختبارات معيارية المرجع. فعندما يكون اهتمام

اختبارات معيارية المرجع منصبا على معرفة انحراف درجة الفـرد عـن متوسط عينة التقنين، ينصب اهتمام الاختبارات محكية المرجع على معرفة القيمة التقديرية لدرجة الفرد في نطاق سلوكي معين عن درجة قطع محددة مسبقا.

وقد اعتمد ليفنجستون في تقدير الثبات على مدى انحراف درجـات الأفراد عن درجة قطع محددة، بدلا من انحراف درجـات الأفـراد عـن متوسط درجات النطاق السلوكي المتبع في حالة اختبارات معيارية المرجع. والمعادلة التالية تعرض معامل ثبات الاختبار محكي المرجع للفنجـستون $(K^2_{(X,T)})$عندما يكون لدينا اختبارين متوازيين:

$$k^2_{(X,T)} = \frac{r_{xx'} s_x s_{x'} + (\bar{x} - nx_0)(\bar{x'} - nx_0)}{\sqrt{[s_x^2 + (\bar{x} - nx_0)^2][s_{x'}^2 + (\bar{x}' - nx_0)^2]}}$$

حيث يشير الحرف X_0 إلى درجة القطع. أما عندما يكون هناك اختبـار واحد فقط فتصبح معادلة ليفنجستون كما هي تاليا:

$$k^2_{(X,T)} = \frac{s_x^2 (KR\ 20) + (\bar{x} - nx_0)^2}{S_x^2 + (\bar{x} - nx_0)^2}$$

وجدير بالذكر أن معامل الثبات المحسوب في هذه الطريقة يتغير بتغيـر قيمة درجة القطع. كما أن المعامل يتأثر بطول الاختبـار. وكـذلك، يتوقـع أن معامل ليفنجستون يطابق معامل كيودر-ريتشاردسون (KR-20) فـي حالـة أن تكون درجة القطع مطابقة لمتوسط درجات الأفراد على الاختبار.

معامل كارفر Carver Index

تعتمد طريقة كارفر على تطبيق اختبارين متوازيين على مجموعة واحدة من الأفراد. ويتم تقدير الثبات بمقارنة نسبة عدد الأفراد الذين يـتم

تصنيفهم على أنهم متمكنين في كل من الاختبارين. فكلما زاد اتساق قـرار التصنيف في مرتي التطبيق كانت الاختبارات أكثر ثباتا. والجدول الثنـائي التالي يعرض الخلايا التي تتضمن أعداد الأفراد حسب تصنيفهم:

		الاختبار (ب)	
		متمكن	غير متمكن
الاختبار (أ)	متمكن	A	B
	غير متمكن	C	D

ويمكن حساب معامل كارفر CI حسب المعادلة التالية:

$$CI = \frac{A + D}{A + B + C + D}$$

وجدير بالذكر أن معامل كارفر يعكس فقط ما إذا كانت نسبة الأفراد لا تتغير في مرتي التطبيق، ولكنه قليل الحساسية لاتساق قرارات تصنيف كل فرد في المرتين.

الثبات واستخدام الاختبار

كثيرا ما يثار تساؤل يتعلق بدرجة الثبات التي يجب أن تتمتع بهـا أداة القياس من أجل أن تكون تلك الأداة مفيدة ؟ وللإجابة على هذا التساؤل، لا بد أن نحدد الأغراض التي سوف تستخدم فيها الدرجات على تلك الأداة. كما أنه ليس هناك من قرار فاعل وأساسي يمكن صنعه على أساس مقياس واحد فقط.

وإذا كانت القرارات التي يراد استخدام الاختبار في صنعها غاية في الأهمية أو لا يمكن إصلاحها فيما بعد, فإن مقدار معامل ثبـات الـدرجات عليه تعد مسألة غاية في الأهمية أكثر مما هو في حالة أن تكون القرارات

عادية أو غير مهمة. من ناحية أخرى، إذا كان سيتم استخدام المقياس للمساعدة في صنع قرارات حول بعض الأفراد، فإنه يجب أن تكون الدرجات أكثر ثباتاً مما لو كانت ستستخدم في صنع قرارات تخص مجموعات.

وإذا توفرت معلومات أخرى قليلة يمكن بناء قرار عليها، فقد يكون من المفيد جداً أن نستخدم اختبارا يتمتع بمستوى ثبات منخفض أفضل من عدم الاختبار (لأن الاختبار ذو الثبات القليل مازال يتمتع ببعض الصدق، وعندها يمكن أن يكون مفيداً).

من زاوية ثانية، إذا كان من الممكن صنع قرار جيد، أو التوصل إلى تنبؤ دقيق دون بيانات تعتمد على الاختبارات، فإن إعطاء الاختبار يعد أمراً غير ذي جدوى علاوة على ما يسببه من عناء، حتى وإن كان ثابتاً.

عند اختيار الاختبارات المقننة، من المهم جداً لقارئ أدلة الاختبارات Test Manual أن يكون قادراً على فهم الثبات وما يتعلق به من بيانات. كما تساعد المعرفة بمفهوم الثبات، والتقديرات المختلفة له، والعوامل المؤثرة على هذه التقديرات في الوصول إلى فهم أعمق.

وتعتمد أنواع بيانات الثبات التي ترد في دليل الاختبار على نوع الاختبار وكيفية استخدامه. ففي اختبارات الاستعداد العام، تكون تقديرات الثبات الأكثر أهمية هي تقدير الاستقرار، لأن نتائج اختبار الاستعداد تستخدم في صنع قرارات تتعلق بتنبؤات طويلة المدى. وعليه، فمن الضروري أن نعرف مدى استقرار نتائج الاستعداد، لأنه إذا لم تكن درجات الاختبار مستقرة، فلن نتمكن من التوقع أو التنبؤ بدرجة دقيقة ومعقولة.

من زاوية أخرى، يبدو من الحيوي أيضاً أن نتعرف على بيانات حول ثبات الاختبارات الفرعية ودرجات الفرق في مقاييس الاستعدادات. لهذا فإن تقديرات التكافؤ والاتساق الداخلي تبدو مفيدة وذات قيمة في تفسير اختبارات الاستعداد، لأن على الفرد أن يمتلك المعلومات الكافية حول تجانس المحتوى ومدى اعتماد الدرجات على بعض الأسئلة المحددة.

في الاختبارات التحصيلية، تبدو تقديرات الثبات أساسية جداً. فقد نحتاج إلى استنتاج مدى إتقان شخص ما للمهارات الأساسية أو المعرفة بمجالات واسعة من خلال استجابات ذلك الشخص على مجموعة محددة من الفقرات. إضافة إلى ذلك، يكون من المفيد أن نمتلك إشارة أو دلالة على مدى تجانس المحتوى. الأمر الذي يستوجب توفير تقديرات ثبات الاتساق الداخلي.

أما بخصوص اختبارات الاستعداد المتنوعة، فإن المفروض أن توفر بطاريات الاختبارات التحصيلية بيانات حول نسب ثبات الاختبارات الفرعية. لأن أغلب الاختبارات التحصيلية مصممة عن قصد لتلائم المنهج الدراسي، ولأن الطلبة يتعلمون هذه المواد بكميات ونسب مختلفة، فانه من غير المتوقع أن تبقى هذه الدرجات ثابتة وعليه، فإن معاملات الاستقرار طويلة الأمد عديمة المعنى نسبياً.

أما في المقاييس غير المعرفية، فإن أنواع ثبات المعلومات المطلوبة تختلف نسبيا. فعلى سبيل المثال، إذا أردنا استخدام اختبار للميل من أجل التنبؤ بالعمل الذي يقتنع به الفرد على المدى البعيد، علينا أن نفترض أساسا بأن الميول ثابتة، وان المعلومات المتعلقة بهذا الافتراض ضرورية جدا (ثبات الاستقرار).

من ناحية أخرى، إذا أردنا الحصول على قياس لسمة وقتية من سمات الشخصية (كالإحباط المؤقت مثلا)، علينا أن لا نتوقع معاملات ذات استقرار عالي، بل علينا أن نبحث عن تقدير ثبات الاتساق الداخلي.

إضافة إلى تقديرات الثبات، يجب أن توفر أدلة الاختبارات بعــض المعلومات الأخرى، كتلك المتعلقة بخصائص العينة التي أعتمد عليها فــي حساب تقديرات الثبات. كحجم تلك العينة وطريقــة اختيارهــا، ومتوســط درجات أفراد العينة على الاختبار وانحرافهــا المعيــاري وغيرهــا مــن مؤشرات وصف توزيع الدرجات

كما أن المفروض أن يتضمن دليــل الاختبــار وصــف الأخطــاء المعيارية في القياس (وكيفية الحصول عليها). كما يجب وصف تقــديرات منفصلة للعمر والمرحلة الدراسية (الصف). كما يجــب عــرض درجــات الخطأ المعياري في القياس ضمن العمر أو المستوى الدراسي الواحد{ على سبيل المثال، يجب أن يذكر اختبار الاستعداد العام أو اختبــار التحــصيل تقديرات الخطأ المعياري للقياس منفصلة لكافة فئات الأفراد المتباينــة ذوي (المستويات العليا، الوسطى، الدنيا).

ملخص

أمكن تلخيص الأفكار التي قدمت في هذا الفصل في النقاط التالية:

- الثبات هو درجة الاتساق بين مقياسين لصفة أو سمة واحدة.
- هناك بعض من مصادر عدم الاتساق أو خطأ التباين مثل عدم استقرار السمة، خطأ المعاينة، خطأ الإجراء أو الإدارة، خطأ التصحيح والأخطاء المتعلقة بالشخص الممتحن.
- إن الخطأ المعياري في القياس هو الانحراف المعياري المقدر لدرجات الفرد الملاحظة عن درجته الحقيقية. فعندما نعرف درجة الفرد الملاحظة، نستطيع أن نحدد فترة ثقة حول هذه الدرجة. ونقول مثلا إننا واثقون حوالي 68 % من أن درجته الحقيقية ستكون ضمن ذلك المدى.
- هناك عدة طرق لتقدير الثبات مثل:(أ) الاستقرار (ب) التكافؤ (ج) الاستقرار والتكافؤ و (د) الاتساق الداخلي.
- يتم الحصول على مقاييس الاستقرار من خلال إجراء اختبار على مجموعة من الأفراد، وإعادة إجراء الاختبار نفسه على الأفراد أنفسهم في وقت لاحق، ومن ثم حساب معامل الارتباط بين مجموعتي الدرجات. وأي تغيير في الدرجة من وقت إلى آخر يعامل معاملة الخطأ.
- يتم الحصول على مقاييس التكافؤ من خلال إعطاء شكلين من الاختبار لمجموعة واحدة من الأفراد في الوقت نفسه ودراسة مدى الارتباط بين الدرجات على الشكلين.

- مقاييس التكافؤ والاستقرار تدمج ما بين أسلوبي الاستقرار والتكافؤ.
- تتطلب جميع تقديرات ثبات الاتساق الداخلي إجـراء واحـد فقـط للاختبار.
- تقوم الطرق المختلفة لتقدير الثبات بالنظر في المـصادر المختلفـة للخطأ ويعتمد تحديد استخدامها على الغرض من اسـتخدام نتـائج الاختبار.
- عموما، تكون الاختبارات الأطول، أكثر ثباتاً. ويتحسن معامل ثبات الاختبار بعد إضافة فقرات ذات خصائص سيكومترية جيدة.
- يجب عدم استخدام تقديرات الاتساق الداخلي في حالـة اختبـارات السرعة.
- يكون الثبات أعلى، عندما يطبق الاختبار على مجموعة متباينة من الأفراد.
- درجات الفرق (أو الاختلاف) أقل ثباتاً من الدرجات المفردة.
- تعتمد تقديرات الثبات التقليدية على تباين الدرجـة الحقيقيـة. ولا ينطبق هذا في حالة الاختبارات محكية المرجع، حيـث لا تكـون التقديرات التقليدية للثبات ملائمة. إذ نعنى بقدرة الدرجة على كشف التباين للدرجة الحقيقية أقل مما نعنى بها في الاختبارات معياريـة المرجع.

أسئلة وتمارين

س1: كل المقاييس عرضة للخطأ. عدد على الأقل خمسة من أخطاء المقاييس المتوقع حدوثها ؟

س2: ما الفرق بين الخطأ العشوائي والخطأ المنتظم ودرجة تأثير كل منهما على تقدير الثبات؟

س3: ما الفرق بين الدرجة الحقيقية والدرجة الخام وما العلاقة التي تربط بينهما؟

س4: ما المقصود بالخطأ المعياري للقياس وما العوامل التي يعتمد عليها في تقدير قيمته؟

س5: لماذا يجب عدم استخدام طريقة التجزئة النصفية في تقدير ثبات اختبارات السرعة.

س6: يذكر أحد أدلة الاختبارات قيمة ثبات تجزئة نصفية مقدارها (0.75)، ماذا يعني ذلك؟ وما قيمة معامل الارتباط بين نصفي ذلك الاختبار؟

س7: حصلت سميرة على الدرجة (75) في اختبار الإملاء. وكان الخطأ المعياري لقياس الدرجات على هذا الاختبار مقداره (3). ما فترة الثقة لدرجة سميرة الحقيقية عندما نكون متأكدين بنسبة (95 %) ؟

س8: إذا كان معامل الارتباط بين نصفي اختبار تحصيلي يتألف من 30 فقرة تم تطبيقه على 50 مفحوصا يساوي (0.75) فما معامل ثبات ذلك الاختبار ؟

س9: إذا كان مقدار تباين نصفي اختبار تحصيلي (4) و(7) على الترتيب. فما مقدار معامل ثبات الاختبار إذا كان تباين الدرجات الكلي للإختبار (18)؟

س10: إذا كان الانحراف المعياري لدرجات (25) طالبا يساوي (3.5) ومعامل ثبات تلك الدرجات (0.64) فأوجد:

أ) الخطأ المعياري لخطأ القياس.

ب) حدي الدرجة (25) إذا أردنا أن نكون واثقين بما نسبتة **95 %** .

س11: اختبار تحصيلي يتألف من 20 سؤالا ويعاني من انخفاض في معامل ثباته (0.55) ، ولذلك أشار أحد العاملين بالقياس النفسي بضرورة أن يغطي الاختبار المحتوى بشكل أفضل بأن يتم إضافة 40 فقرة جديدة وفعالة للاختبار. كم سيصبح معامل ثبات الاختبار الجديد ؟

س12: أراد معلم أن يحسب معامل ثبات اختبار يتألف من 30 فقرة موضوعية تم تطبيقه على 40 طالبا وجمع البيانات التالية:

$\bar{x} = 22$ $S_x = 3,$ $\sum pq = 3,$

أحسب معامل ثبات الاتساق الداخلي:

أ) باستخدام معادلة KR-20

ب) باستخدام معادلة KR-21

س13: اختبار تحصيلي يتألف من 80 سؤالا ويمتاز بارتفاع في معامل ثباته (0.90) ، وبالرغم من ذلك، أشار أحد العاملين بالقياس النفسي بضرورة أن يتضمن الاختبار 40 سؤالا فقط بسبب مقدار الوقت المخصص للاختبار. فكم سيصبح معامل ثبات الاختبار الجديد؟

س14: اختبار تحصيلي يتألف من 90 سؤالا ويمتاز بمعامل ثباته المرتفع (0.90) ، وبرغم ذلك، أشار أحد العاملين بالقياس النفسي بضرورة أن يتضمن الاختبار 30 سؤالا فقط بسبب الوقت المخصص للاختبار. فكم سيصبح معامل ثبات الاختبار الجديد ؟

س15: ما علاقة ثبات الاختبار بمعامل صعوبة فقراته؟

س16: كيف يتغير معامل ثبات الاختبار حسب عدد فقراته ؟

- **أجب بـ "نعم" أو بـ "لا" على العبارات التالية:**

*. يزداد الخطأ المعياري للقياس كلما يزداد الانحراف المعياري لـدرجات المفحوصين.

*. يزداد معامل ثبات الاختبار كلما قلت درجة تجانس المفحوصين.

*. إذا طبق الاختبار على عينة من الطلبة المتفوقين فسيكون معامل الثبات عاليا.

*. تكون الدرجة الحقيقية أقل من الدرجة الخام.

*. يعمل الضجيج بجوار قاعة الامتحان على إحداث خطأ عشوائي في الدرجات.

*. يعمل الخطأ المنتظم في درجات المفحوصين على إنقاص معامل ثبات الاختبار.

*. يزداد معامل ثبات الاختبار حتما بزيادة عدد فقراته.

*. تفترض معادلتا كودر-ريتشاردسون أن جميع فقرات الاختبار متساوية في الصعوبة.

*. معادلة جوتمان هي حالة خاصة من معادلة كرونباخ ألفا.

*. تعد معادلة كرونباخ ألفا حالة خاصة من معادلة سبيرمان-براون في حساب ثبات التجزئة النصفية.

*. يصل معامل ثبات الاختبار الكلي أقصى قيمة له إذا كانت العلاقة بين نصفيه تامة.

*. يصنف الخطأ العيني ضمن الأخطاء التي تتعلق بظروف تطبيق الاختبار.

*. الخطأ العشوائي هو الخطأ الذي نهتم به عند تقييم معامل ثبات الاختبار.

*. يزداد معامل ثبات الاختبار كلما اقترب معامل الارتباط بين نصفيه من الصفر.

*. إن إضافة فقرات سهلة جدا إلى الاختبار، يزيد مقدار معامل ثباته.

*. لا يؤثر إضافة فقرات صعبة جدا إلى الاختبار على معامل ثباته.

*. يتمتع الاختبار الذي يتألف من فقرات انتقائية بثبات نسبي أكبر من ذلك الذي يتألف من فقرات صياغية.

*. عموما يتمتع معامل ثبات الفرق بين الدرجات على اختبارين بمقدار يقل عن ثبات أي منهما.

*. يزداد معامل ثبات الفرق بين الدرجتين كلما قل معامل الارتباط بين درجات الاختبارين.

*. تعطي طريقة ثبات الاستقرار والتكافؤ تقديرا أقل مما تعطيه أي من طريقتي ثبات الاستقرار أو ثبات التكافؤ.

- **ضع دائرة حول رمز الإجابة الصحيحة فيما يلي:**

*. إذا كان معامل الارتباط بين درجات الفقرات الفردية ودرجات الفقرات الزوجية لإختبار يساوي (70.) فإن ثبات الاختبار كاملا يساوي:

أ) 70. ب) 82.

ج) 80. د) المعلومات غير كافية

*. يشير ثبات الاختبار إلى ---.

أ) استقرار النتائج ب) تجانس محتوى الاختبار

ج) موضوعية تطبيق الاختبار وتصحيحه د) صدق الاختبار

*. إذا تم ضبط كافة العوامل المؤثرة في ثبات الاختبار، أي أنواع فقرات الاختبار تعطي أقل معامل ثبات؟

أ) فقرات الإكمال ب) الفقرات الإنشائية

ج) الصواب والخطأ د) فقرات المزاوجة

*. إذا كان مقدار تباين الجزء الخطأ في درجات المفحوصين يساوي (30 %) من التباين الكلي للاختبار، فإن معامل ثبات الاختبار يساوي----.

أ) 30. ب) 70.

ج) 60. د) البيانات المعطاة غير كافية

*. إذا كان معامل الارتباط بين درجات الفقرات الفردية ودرجات الفقرات الزوجية لإختبار يساوي (1) فإن ثبات الاختبار كاملا يساوي:

أ) 5. ب) 1

ج) صفر د) المعلومات المعطاة غير كافية

*. إذا كان معامل ثبات كل اختبار يساوي (1)، فإن ثبات الفرق بين الدرجات على الاختبارين يساوي ----.

أ) 1 ب) صفر

ج) 5. د) المعلومات المعطاة غير كافية

الفصل السادس

الخصائص السيكومترية لأدوات القياس

ثانيا : الصدق Validity

- مقدمة
- علاقة صدق الإختبار بثباته
- أنواع الصدق
    - صدق المحتوى
    - الصدق المرتبط بمحك
    - صدق البناء أو التكوين (المفهوم)
        - الصدق التقاربي
        - الصدق التمايزي
        - الصدق العاملي
- الصدق المتقاطع (عبر العينات)
- التداخل بين مؤشرات الصدق
- صدق الإختبارات محكية المرجع
- ملخص
- أسئلة وتمارين

6

المخرجات التعليمية

بعد دراسة هذا الفصل يتوقع من الطالب أن يكون قادرا على أن

- يعرف مفهوم صدق أداة القياس وعلاقته بالثبات.
- يحدد الهدف من إجراء صدق المحتوى بمختلف أنواعه.
- يقارن بين مفهومي صدق المحك التنبؤي وصدق المحك التزامني.
- يحسب معامل الصدق المرتبط بالمحك من خلال دقة القرار وجداول التوقع.
- يتعرف على أساليب صدق البناء (التحليل المنطقي، الفروق بين المجموعات الإرتباطية، التجريبية، التحليل العاملي).

مقدمة

يعد صدق أداة القياس من الخصائص السيكومترية الأكثر أهمية مقارنة مع الخصائص الأخرى كالثبات، وذلك بسبب ارتباط الصدق بالهدف أو بالأهداف المتوقع من أداة القياس تحقيقها، وكذلك بمدى اتصاله بنوع وأهمية القرار الذي سيتم اتخاذه تبعا لذلك. ويعرف صدق الأداة بأنه المدى الذي تقيس تلك الأداة ما بنيت من أجله.

وعندما نتحدث عن صدق أداة القياس، فإننا نشير إلى أداة فاعلة وصالحة لتحقيق أهداف معينة، ترتبط بمحتوى دراسي أو قدرة عقلية، أو بمجال يتعلق بقياس سمة نفسية، وموجهة إلى فئة معينة من الأشخاص. وهذا يفيد بأن الصدق مفهوم نسبي relative وليس مطلقا absolute، إذ لا يجوز القول بأن اختبارا ما صادق لتحقيق كل الأهداف ويفي بكل الأغراض ويستخدم مع كافة المستهدفين من كل الفئات العمرية.

من جهة ثانية، فإن الصدق لا يرتبط بأداة القياس ذاتها، بل بطريقة تفسير الدرجات المستخرجة من تلك الأداة. فقد يتم استخدام اختبار في القدرة العددية مثلا في أكثر من مجال، كأن تستخدم النتائج على هذا الإختبار في **عمليات قبول** طلبة في برنامج معين، أوفي **تشخيص** قدرات أشخاص معينين، أو ربما في **التنبؤ** بمستوى أداء مجموعة من المتقدمين لشغل وظيفة ما. واختبار القدرة العددية الصادق يزودنا بدرجات يمكن استخدامها لتصنيف الأفراد حسب قدرتهم العددية فعلا. بينما يعمل مقياس الشخصية الصادق على إعطاء درجات على هذا المقياس بحيث تبين فروقا ذات معنى في الشخصية لدى الأفراد الذين صمم المقياس من أجلهم (Crocker and Algina, 1986).

من هنا يتضح أن صدق أداة القياس يأخذ تنوعا من المعاني بناءا على الغرض الذي يبنى من أجله في كل مرة. فعند الحديث عن صدق أداة

القياس، يتجه الاهتمام حتما نحو صلاحية تلك الأداة في تحقيق غرض معين أو أغراض معينة، وليس نحو الأداة نفسها (Sax, 1981). وبنفس السياق، يرى عدد من مشاهير العاملين في مجالات القياس في العلوم السلوكية (Anastasi, 1976; Cronbach, 1970; Nunnally, 1978; Stanley, 1981) أن الصدق هو مجموعة الأدلة evidences التي نسترشد بها للتحقق من وجود القدرة أو السمة، أو حتى السلوك.

وبالرغم من الفكرة التي أطلقها ساموبل ميسيك (Messick, 1989) حول المفهوم الموحد unitary concept للصدق، إلا أن تصنيف الصدق في ثلاث أنواع هي صدق المحتوى والصدق المرتبط بمحك وصدق البناء أو التكوين ما زال شائعا ومتبعا في معظم كتب القياس. وقد أشار ميسيك في بحثه إلى أن الصدق هو الدرجة التي تعمل الأدلة العملية والمبررات والأطر النظرية في دعم مدى الدقة والملاءمة للتفسيرات والإجراءات والقرارات التي تعتمد على الدرجات التي نحصل عليها من خلال استخدام أداة القياس.

ومعروف أن إجراءات صدق أداة القياس تبدأ منذ مراحل بناء تلك الأداة. إذ يتم الوصول إلى صدق الإختبار كمفهوم شامل من خلال تفحص كل خطوة من خطوات اختيار أو بناء ذلك الإختبار. فالتأكد من الأهداف التي يوضع الإختبار لقياسها، من حيث الصياغة السليمة لتلك الأهداف من ناحية، وتحديد الأهمية النسبية لكل منها، ومدى مناسبتها لمستوى المستهدفين من ناحية أخرى، يعد بداية رئيسة في الحكم على صدق الإختبار. إضافة إلى دقة تحديد المجال السلوكي الذي يتوقع أن تمثله أداة القياس، وكذا الصياغة الدقيقة لمكوناته وعباراته.

علاقة صدق الإختبار بثباته

عرفنا في الفصل السابق أن معامل الثبات إحصائيا هو نسبة التبـاين الملاحظ الذي يعزى إلى التباين الحقيقي. بينما يعـرف معامـل الـصدق إحصائيا بمقدار نسبة التباين الحقيقي العائد للسمة بالنسبة للتباين الملاحظ. والمعروف أن التباين الحقيقي S_T^2 يتألف من نوعين من التباين هما:

أ) تباين حقيقي عائد للسمة المقاسة S_R^2

ب) تباين حقيقي غير عائد للسمة المقاسة S_I^2

$$S_T^2 = S_R^2 + S_I^2$$

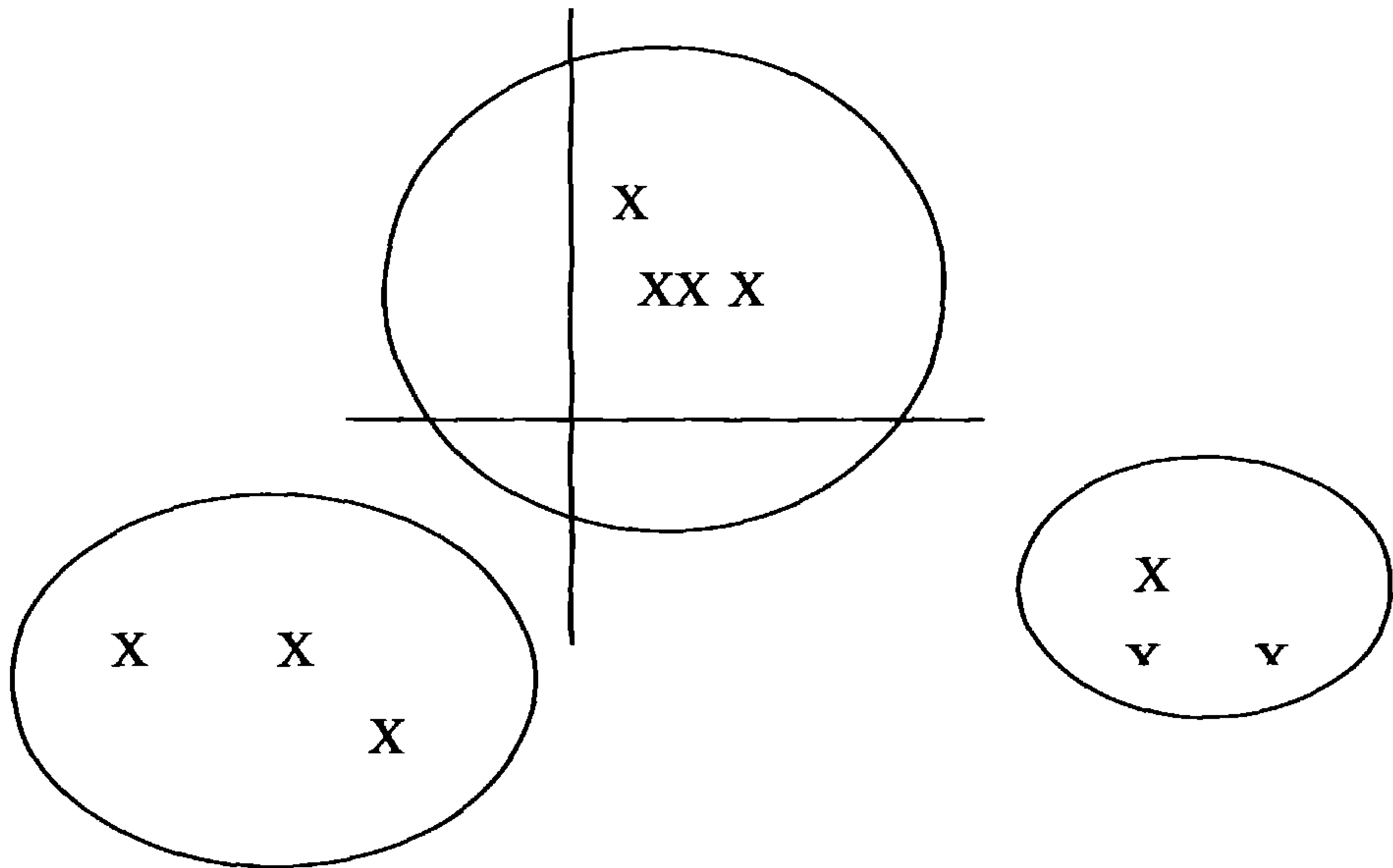
فإذا كانت السمة المنوي قياسها هي القدرة العددية لدى أطفال المرحلة الإعدادية، فإن التباين بين أداء المفحوصين على حل المسائل وفهم المفاهيم ذات العلاقة بالقدرة العددية ومكوناتها يعتبر تباينا حقيقيا ويقع في صـميم السمة المقاسة (S_R^2). بينما قد يستوجب الأداء على مسائل القدرة العددية قدرة لغوية وقدرة على الاستيعاب، فهذه قدرات حقيقية ولكنها لا تنسب إلى السمة المنوي قياسها (S_I^2)

وكذلك لو كانت السمة المنوي قياسها هي القدرة على استيعاب مفاهيم الديناميكا الحرارية في الفيزياء، فسيكون التباين في الدرجات على الفقرات التي تمثل هذه القدرة حقيقيا وعائدا إلى السمة (S_R^2). بينما قـد يتفـاوت المفحوصون في قدراتهم على حل مسائل الديناميكا الحرارية حسب تمكنهم من بعض المعالجات الرياضية التي هي أيضا تفرز تباينا حقيقيا ولكنه غير عائد إلى السمة (S_I^2)

من هنا يتضح أن معامل الثبات أشمل من معامل الصدق. بمعنى أنه إذا زاد معامل الصدق فإن معامل الثبات سيزداد حتما، أما إذا زاد معامـل الثبات فقد يزداد معامل الصدق وقد لا يزداد. لذلك اشتهر القول بأن الثبات

شرط لازم للصدق ولكنه ليس كاف. بمعنى أن زيادة معامل الثبات قد تكون بسبب التحسن في الجزء الحقيقي من التباين العائد إلى السمة موضع القياس.

من زاوية أخرى، يشير الصدق Validity إلى صحة القياس Accuracy ، بينما يشير الثبات Reliability إلى دقته Precision. ويبين المخطط التالي وصفا لمفهومي الصدق والثبات. فالقياسات في الوضع (A) هي البيانات التي تتسم بالصدق والثبات. بينما تشير البيانات في الوضع (B) إلى غياب كل من الصدق والثبات، حيث تظهر عدم دقة القياسات بسبب تباعدها بشكل واضح (عدم الدقة)، واختلافها الواضح عن الوضع (A). أما الوضع (C)، فيشير إلى بيانات تتسم بالثبات (دقة القياس) لتقارب قيمها، بالرغم من أنها غير صادقة، لأنها تقع جميعها خارج الهدف المنشود والموضح بالوضع (A):



أنواع الصدق

جاء تصنيف أنواع الصدق في ثلاث أنـواع رئيـسة هـي صـدق المحتوى، والصدق المرتبط بالمحك، وصدق البناء (التكوين) (Nunnally, 1978). وسيتم التعرض لكل منها بنوع من التفصيل:

صدق المحتوى Content Validity

يتم دراسة صدق المحتوى من خلال عملية التحليل المنطقي لمحتوى أداة القياس اعتمادا على حكم فـردي (Allen and Yen, 1979). ولأن الحكم الفردي يعد جهدا ذاتيا، فإنه ربما يفتقر إلى الموضوعية. الأمر، الذي ربما يزيد من احتمالية الوقوع بخطأ التقدير. وبشكل عام، يأتي الحديث عن صدق المحتوى عندما تتجه نية مستخدم أو مطور أداة القياس إلى الوصول إلى قرار حول مجال واسع من الفقرات تشبه ما تتضمنه تلك الأداة، اعتمادا على درجات الذين طبقت عليهم الأداة (Crocker and Algina, 1986).

وترتبط دراسات صدق المحتوى بشكل كبير بالإختبارات التحصيلية بأنواعها المقننة أو غير المقننة التي تصمم من قبل شخص مهتم كمعلـم أو أخصائي في مهنة ما، وكذلك اختبارات برامج التدريب بأنواعها. إضـافة إلى كافة الإختبارات المستخدمة للحكم على مستوى أداء الطلبة في نهايـة أي مرحلة دراسية.

من زاوية أخرى، يرتبط صدق المحتوى بأدوات القياس الأخرى غير التحصيلية (المقاييس النفسية والاجتماعية والشخصية وغيرها). فبدلا مـن المحتوى الدراسي والمنهاج في حالة الإختبـارات التحـصيلية والقـدرات الأكاديمية، يتم التركيز على المجال السلوكي الذي يعرف السمة النفسية أو القدرة العقلية والاستعدادات وغيرها.

ويتم الحكم على درجة صدق المحتوى للإختبارات التحصيلية من مدى تمثيل فقرات الإختبار لجدول المواصفات. ولأن جدول المواصفات ينبثق من توجهات باني الإختبار وأهدافه، بات من الضروري استعراض نوعين من الصدق يتصلان بصدق المحتوى هما الصدق الظاهري والصدق المنطقي أو العيني (Allen and Yen, 1979):

- الصدق الشكلي أو الظاهري face validity ويعكس مدى انسجام فقرات الإختبار مع موضوع الإختبار ومفاهيمه وفقا لتعريف باني أو مطور الإختبار.
- الصدق المنطقي أو العيني logical or sampling validity ويرتبط بتصنيف الفقرات وفقا لمجالات أداة القياس أو الأبعاد المكونة لأدوات القياس.

ويتم درس الصدق الشكلي من خلال التبصر في مضمون فقرات الإختبار والحكم على مدى علاقة فقرات الإختبار بمحتوى المادة الدراسية (عودة، 2000). وخير مثال على ذلك، اختبارات المواد الدراسية التي يحدث فيها تداخل مع مواد دراسية أخرى كمادة الفيزياء التي تتداخل كثيرا مع الرياضيات واللغة، أو اختبار في الحساب للصف الأول الابتدائي عندما يتألف من فقرات يحتاج حلها قدرة لغوية في القراءة والاستيعاب. عند ذلك فإن اختبار الحساب سيفتقر إلى صدقه الشكلي، كون الإجابة على فقرات ذلك الإختبار لا تستوجب قدرة حسابية فقط بل وقدرة لغوية بدرجة معينة كذلك. وبنفس الوقت لن يكون اختبار الفيزياء للمرحلة الثانوية صادقا بدرجة كبيرة إذا احتوى فقرات يحتاج حلها إلى قدرات رياضية بمستوى ما ومهارات لغوية في الاستيعاب والتحليل بدرجة أخرى أيضا (Crocker and Algina, 1986).

أما الهدف من دراسة الصدق المنطقي أو العيني أو المحتوى لاختبار ما يتلخص في التأكد فيما إذا كانت فقرات الإختبار تمثل مجال أداء (منهاج أو محتوى دراسي) يتعلق بسمة أو قدرة. ففي اختبار مفردات لغوية مثلا، لا يكون اهتمام مطور الإختبار التعرف فقط على ما إذا كان الطلبة يعرفون فقط معاني تلك المفردات، بل سيكون اهتمامه في تقويم معرفة الممتحنين في مفردات تشبه تلك التي احتواها الإختبار، أو مدى استيعابه للمجال السلوكي الذي يتضمن المفردات اللغوية التي تشبه تلك التي وردت في الإختبار (Magnusson, 1967).

إن عملية الحكم على صدق محتوى أداة قياس معينة يعد عملا منطقيا وتحكيميا ينفذه فريق من الخبراء والمهتمين والمستقلين في مجال أداة القياس. ويتحدد هدف فريق الخبراء في عملية الحكم على مدى تمثيل فقرات الإختبار للمحتوى الذي اشتقت منه، ويراعي ذلك عوامل الشمول والتنوع والدقة. وجدير بالذكر أن إجراءات صدق المحتوى الرئيس تأتي منذ التكوين الأولي لفقرات أداة القياس، ويقوم بتنفيذ هذه الأنشطة والإشراف عليها مطور الإختبار، أو الجهة المستفيدة منه، مستبعدا هؤلاء الذين قاموا بكتابة وصياغة الفقرات. وبشكل عام، فإن دراسات صدق المحتوى تتضمن الخطوات التالية بحدها الأدنى:

- تحديد مجال الأداء موضع البحث (مفردات لغوية لمستوى الصف السادس الابتدائي).
- بناء أداة القياس (تحليل وحدات المجال السلوكي الذي تقيسه الأداة، وتحديد أهمية كل منها).
- اختيار فريق من الخبراء المحترفين Professional Experts في مجال الأداء أو المحتوى الدراسي الذي تغطيه أداة القياس ومن يدرس الفئة العمرية أو دونها.

- بناء خطة مفصلة لطريقة مطابقة الفقرات مع المجال السلوكي الذي تم تحديده.
- جمع وتلخيص البيانات والمعلومات من عملية المطابقة آنفة الذكر.

من هنا تبرز أهمية التعرف على الأهداف التدريسية Instructional objectives كما جاءت في تصنيف بلوم (Bloom, 1965) من ناحية، والإلمام بطريقة بناء جداول المواصفات التي ورد الحديث عنها في الفصل الثاني من هذا الكتاب من ناحية أخرى. ويقترح كروكر وألجاينا (Crocker and Algina, 1986) ضرورة الاهتمام بالاعتبارات التالية عند دراسة صدق المحتوى:

1. **إلى أي مدى كانت أوزان الأهداف التدريسية تعكس أهميتها الفعلية؟**

   حيث تختلف الأهداف في أهمية كل منها، مبينا ضرورة تحديد أوزانها بشكل مسبق، ويتم هذا من قبل خبراء في تدريس وتقييم المادة الدراسية لموضوع الإختبار. ويمكن أن يقوم بتلك المهمة مدرسو المادة العلمية في وزارة التربية والتعليم ذوو العلاقة، مع ضرورة أن يتم توضيح مفهوم الأهمية النسبية من قبل مطور الإختبار، وليس شرطا أن يكونوا هم (أي مدرسو المادة) الذين قاموا بكتابة الفقرات.

2. **كيف يمكن تصميم عملية مطابقة matching فقرات أداة القياس؟**

   فقد يطلب من كل خبير أن يفحص مدى مطابقة كل فقرة لقائمة الأهداف. ويقترح كل من كاتز (Katz, 1958) وإيبل (Ebel, 1979) بضرورة قراءة الفقرة بإمعان وتحديد الإجابة الصحيحة كما يقوم بذلك المفحوص. أما كلاين وكوزيكوف (Klein and

Kosecoff, 1975) فقد اقترحا ثلاثة خطوات إضافية لتسهيل عملية المطابقة يقوم بها المحكمون، وهي:

- كتابة كل فقرة على بطاقة منفصلة،
- مقارنة كل فقرة مع الأهداف التي يتوقع أنها تقيس أحدها أو بعضا منها،
- مقارنة المطابقة والحكم على مدى مطابقة الفقرة للهدف، وذلك بتفريغها في جدول مخصص لهذا الغرض.

ويعتقد البعض أن المطابقة عملية ثنائية البعد، أي أن تكون الفقرة مطابقة للهدف أم غير مطابقة، بينما وصف هامبلتون (Hambleton, 1980) عملية المطابقة بأنها ليست بالضرورة أن تكون ثنائية، بل يمكن اتباع سلم تقدير يتألف من خمسة مستويات يتراوح بين (1) عديمة المطابقة إلى (5) مطابقة تماما. كما يمكن حساب المتوسط أو الوسيط لمجمل التقديرات التي يعطيها المحكمون لمدى مطابقة الفقرة الواحدة للهدف.

3. **ما جوانب الفقرة التي يجب فحصها؟** إذ من المفروض تزويد المحكمين بوصف واضح للفقرة والمجال السلوكي الذي تمثله الفقرة، إضافة إلى المادة الدراسية، والعمليات العقلية، ومستوى تعقيد الأداء المطلوب للإجابة، وكذلك نمط أو شكل الإجابة. وإن التأمل في الشكل التالي ربما يوضح شيئا من هذا:

تأمل الهدفين التاليين اللذين يقيس كل منهما القدرة الحسابية لدى طلبة الصف الأول الابتدائي:

أ. يجمع عددين موجبين بحيث يكون مجموعهما 18 أو أقل.
ب. يطرح عددين كل منهما أقل من 20 بحيث يكون الناتج موجبا.

وقد تم بناء الفقرات الست التالية لقياس هذين الهدفين:

1) 3 + 5 = ——
2) 12 - 10 = ——
3) 8 - 5 = ——
4) 25 - 16 = ----
5) 13 + 3 - 8 = ——

6) إذا كان مع سعيد 10 دولارات وخسر دولارين ، ف كم بقي معه ؟

يلاحظ مما تقدم أن الفقرة رقم (1) تطابق الهدف (أ)، بينما تطابق الفقرتان (2) و (3) الهدف (ب)، وتعالج الفقرة الرابعة مادة مختلفة لم يتطرق لها أي من الهدفين ، بينما تتضمن الفقرة الخامسة نفس المادة التعليمية المتضمنة بالهدفين ولكنها تتطلب مستوى أعلى من الأداء المطلوب في الهدفين منفردين. أما الفقرة الأخيرة رقم (6) فلا تحوي فكرة الجمع والطرح البسيط فقط، بل إن حلها يستوجب قدرة على القراءة والاستيعاب أيضا وهي تقيس الهدف (ب). من هنا يمكن ربط هذه الفكرة بمفهوم الصدق الشكلي للإختبار.

وحول شكل الإجابة، من المفروض أن يتم عـرض المـسألة فـي الإمتحان بالشكل الذي تم عرضه في المحتوى أو كيفما جرى تدريسه على السبورة في الصف. فالفقرات من 1 – 5 تم عرضها بشكل أفقي، ويخشى أن يتغير أداء الطلبة على كل منها، فيما لو عرضت في الإمتحان بـشكل عمودي.

4. **كيف يمكن تلخيص نتائج وملاحظات الفحص؟** فقد يمكن القول هنا بأن عملية الحكم فيما إذا كانت فقرة أو عدد من الفقرات تطـابق هـدفا أو مجالا أو محتوى ما هي قرار ذو طبيعة نوعية وليس مـسألة كميـة. ومهما يكن من أمر، فإن المؤشرات الكميـة التـي تلخـص قـرارات المحكمين تعد أمرا غاية في الأهمية وهي على النحو التالي (Crocker and Algina, 1986):

- نسبة الفقرات التي طابقت الهدف الذي تنتمي إليه.
- نسبة الفقرات التي طابقت أهدافا وفق مستوى يتدرج من 1 (غير مطابقة) إلى 5 (مطابقة تماما).
- ارتباط الأهمية النسبية للأهداف بعدد الفقرات التي تقـيس تلـك الأهداف.
- نسبة الأهداف التي لم يتم قياسها من أي من فقرات الإختبار.

والجدير بالذكر أن النتائج المعتمدة على تلك المؤشرات لا تؤدي بالضرورة إلى نتائج متماثلة نظرا لاعتماد كل منها على تقدير أو مبرر منطقي معين. يعكس درجة عالية من ذاتية الخبير، كما أن هناك اعتبارات وافتراضـات يستوجبها كل مؤشر. فالمؤشران الأول والثاني يتطلب عددا كبيرا نسبيا من الفقرات (100 فقرة أو أكثر). ويعتمد المؤشر الثالث على التباين في عـدد الفقرات التي تقيس كل هدف ونسبة أهمية كل هدف. ويعـرض الجـدول التالي الحد الأدنى لنسبة قبول العبارة أو الفقرة عند تحكيمها:

عدد الخبراء	عدد الخبراء (المحكمين) الذين أيدوا تمتع الفقرة بصدق محتوى								
	2	3	4	5	6	7	8	9	10
2	1								
3	0.67	1							
4	0.50	0.75	1						
5	0.40	0.60	0.80	1					
6	0.33	0.50	0.67	.83	1				
7	0.29	0.43	0.57	0.71	.85	1			
8	0.25	0.38	0.50	0.63	0.75	.88	1		
9	0.22	0.33	0.44	0.56	0.67	.78	.89	1	
10	0.20	0.30	0.40	0.50	0.60	0.70	.80	.90	1

يلاحظ من البيانات الواردة في هذا الجدول أنه عندما يكون المحكمون أربع أو أقل، فإن الفقرة تكون مقبولة إذا حظيت بموافقتهم جميعا. ويبين الخط الغامق للنسب المقبولة لموافقات المحكمين على العبارات.

من ناحية أخرى يرتبط بصدق المحتوى مفهومان يتعلقان بالمنهاج أو المادة الدراسية هما:

- الصدق المنهاجي Curricular Validity
- الصدق التدريسي Instructional Validity

ويقصد بالصدق المنهاجي مدى تمثيل الفقرات لأهداف منهاج دراسي معين كما تم وصفه آنفا. بينما يشير الصدق التدريسي إلى مدى تمثيل فقرات الإختبار في قياس أهداف جاءت في الدروس والمحاضرات التي تم إلقاؤها على الطلبة (McClung, 1978). وقد جاء البحث في هذين النوعين من الصدق وتمييزهما عن صدق المحتوى جراء القضايا التي عالجتها المحاكم في الولايات المتحدة التي أمرت بمنح درجات علمية بناءا على اجتياز الطلبة اختبارات كانت تقيس مدى دراستهم لمحتوى دراسي معين

بصرف النظر عن مدى العلاقة التفصيلية بما تم دراسته في منهاج أو محتوى دراسي ما (Yallow and Popham, 1983).

وثمة مسألتان غاية في الأهمية يتعلق كل منهما بصدق محتوى أداة القياس، **الأولى** كأن تكون الفقرة مطابقة لهدف دراسي معين بدرجة كبيرة ولكن الهدف نفسه قد لا يكون مطابقا للمحتوى الدراسي موضع البحث، لكنه يتمتع بأهمية معينة، فعملية مطابقة الفقرة بالهدف لا تقل أهمية عن مطابقة الهدف بالمحتوى الدراسي. **والمسألة الثانية** هي تحيز المحتوى Content Bias كأن يركز الإختبار على قياس مستويات أهداف معينة دون أخرى أو بدرجات اهتمام متفاوتة، أو قد يهتم الإختبار بوحدات دراسية دون أخرى وبدرجات متفاوتة كذلك.

من ناحية أخرى قد يظهر هناك تحيز Bias في بعض الإختبارات التي تهتم بمحتويات دراسية ثقافية أو تاريخية. أو قد يظهر تحيز ضد متحدثي لغة أو لغات معينة، كأن يكون هناك اختبار في الرياضيات وقد كتبت فقراته باللغة الفرنسية وموجه لطلبة المرحلة الإعدادية من جنسيات متعددة. عند ذلك، سيكون أداء الطلبة الذين يتكلمون الفرنسية أفضل بكثير من أقرانهم ممن يتكلمون لغات غير الفرنسية. وجدير بالذكر أن ثبات الاتساق الداخلي بين فقرات الإختبار التحصيلي يشير إلى أن تلك الفقرات تميل إلى قياس شيء مشترك وهذا يدعم صدق الإختبار، بصرف النظر عن مستوى أو نوعية الفقرات. ويشبه ذلك إلى حد كبير ارتباط الدرجات على اختبارات مختلفة. وهذا يدعم أيضا مقولة أن هذه المقاييس تقيس شيئا واحدا حتى لو كان خطأ.

ثمة نوع آخر من الأدلة الظرفية Circumstantial Evidence لصدق محتوى الإختبار، عندما يتم مقارنة الأداء على ذلك الإختبار قبل وبعد فترة تدريب أو تدريس معينة، فيكون الإختبار صادقا إذا حصل تحسن جوهري في درجات الممتحنين بعد تلك الفترة.

الصدق المرتبط بمحك Criterion-Related-Validity

تتجه النية أحيانا لتطوير أداة قياس تطبق على مجموعة من المستهدفين بوقت معين، بقصد استخدامها لأغرض التنبؤ بأداء تلك المجموعة في وقت ما في المستقبل. في هذه الحالة، فإن إجراءات الصدق المتعلقة بتلك الأداة تنصب حول ما إذا كانت أداة القياس قادرة على توفير الفرصة للتنبؤ بالأداء المستقبلي بدرجة تتمتع بدلالة كافية.

وخير مثال على تلك الأداة، اختبارات القبول المستخدمة لأغـراض قبول الطلبة في الدراسة الجامعيـة أو الإختبـارات المـستخدمة لترشـيح مجموعة من المتقدمين لشغل وظيفة أو مهنة، أو الاشتراك فـي برنـامج تدريبي أو إرشادي معين. وتلعب هذه الأداة دور المتنبىء predictor بينما يعرف مستوى الأداء أو السلوك الذي يتوقع من الأداة التنبؤ بـه بالمحـك criterion (Crocker and Algina, 1986).

يدرس الصدق المرتبط بالمحك من خلال تقييم درجة الإرتباط بـين المتغيرات. ويعرف معامل الإرتباط بين الدرجات على المتنبىء والدرجات على المحك بمعامل الصدق المرتبط بالمحك، ويرمز له بالرمز r_{xy} حيـث يرمز الحرف x إلى المتنبىء ويشير الحرف y إلى المحك. ويقع الـصدق المرتبط بمحك في نوعين هما:

- الصدق التنبؤي Predictive Validity
- الصدق التزامني Concurrent Validity

وفي كلا النوعين من الصدق المرتبط بالمحك، يـتم تطـوير أداة القيـاس (المتنبىء) ومن ثم تطبيقه على الأشخاص المعنيين ثم الانتظـار لفتـرة زمنية معينة، ويتم جمع درجات هؤلاء الأشخاص على المحك. ثم يحـسب معامل الإرتباط بين درجات الأشخاص في المرتين وهـذا هـو معامـل الصدق المرتبط بالمحك لأداة القياس التي تم تطويرها آنفا.

والفترة الزمنية المذكورة هي التي تحدد نوع الصدق تنبؤيا كان أو تزامنيا. إذ تكون الفترة في حالة صدق التنبؤ طويلة نسبيا (فصلا دراسيا أو مدة برنامج التدريب، أو طيلة فترة البرنامج الإرشادي)، في حين يتم تطبيق المتنبىء والمحك على المعنيين في نفس الوقت وبدون فاصل زمني في حالة الصدق التزامني. وفي المستطيل التالي مثال على معامل صدق التنبؤ:

معامل الإرتباط بين امتحان الثانوية العامة مقدرا بمعدلات الطلبة في هذا الإمتحان، والنجاح مستقبلا بالدراسة في الجامعة مقدرا بالمعدلات التراكمية الجامعية، هو معامل الصدق التنبؤي لامتحان الثانوية العامة

يلاحظ هنا أن فترة زمنية طويلة تفصل بين الانتهاء من الثانوية العامة والانتهاء من الدراسة الجامعية أو على الأقل بعد مضي سنوات دراسية جامعية معينة. بينما يمثل المستطيل التالي نموذجا للصدق التزامني:

معامل الإرتباط بين تشخيصات المتدربين النفسيين لمراجعي العيادات النفسية، وتشخيصات الإخصائيين المرخصين للغرض نفسه، هو معامل الصدق التزامني لتشخيصات المتدربين.

ويلاحظ هنا أن القياسين يحدثان في زمن واحد ولا يوجد فاصل زمني بينهما. ومهما يكن من أمر، فإن إجراءات الصدق التنبؤي تتلخص بالخطوات التالية كما وردت في كروكر وألجاينا(Crocker and Algina, 1986):

1. تحديد السلوك المحكي المناسب وطريقة قياسه.
2. تحديد عينة من الممتحنين الذين يمثلون من سيتم استخدام الإختبار لصالحهم.
3. تطبيق الإختبار (المتنبىء) على أفراد العينة والاحتفاظ بسجل درجات لكل فرد.

4. عندما تتوفر بيانات عن المحك، يتم جمع بيانات عن أداء كل فرد على المحك.
5. تقدير قوة العلاقة بين الدرجات على الإختبار (المتنبىء) والدرجات على المحك.

وإذا كان الصدق التنبؤي يهتم بالتنبؤ، فإن الصدق التزامني يهتم بالوصف. ويستخدم الصدق التزامني كثيرا في مجالات الصناعة وانتقاء الأفراد والتشخيص الاكلينيكي، فمثلا يمكن القيام تحليل العمل Job analysis لتحديد المهارات أو السلوك الذي يتطلبه ذلك العمل، وتقدير درجة امتلاك الفرد الذي يود الالتحاق بهذا العمل لعينات من السلوك ممثلة لهذه المهارات (علام، 2000).

يفيد الصدق التزامني في تقدير الصدق التنبؤي للإختبارات الصفية، فالمعلم عادة ما يهتم بالسلوك المستقبلي لتلاميذه، ومن الصعوبة عليه إجراء دراسات تتبعية لتقدير مدى صدق الأساليب التعليمية التي يتبعها، لذلك يمكنه تقدير صدق هذه الأساليب من خلال سلوك تلاميذه في الوقت الراهن. إذ من المعلوم أن العلاقة بين متغيرين تقل مع مرور فترة زمنية طويلة نسبيا (Glass and Hopkins, 1996). وعليه فإن اختبارات الاستعدادات ترتبط بالتحصيل الدراسي في الوقت الراهن بدرجة أكبر من ارتباطها به بعد مرور عدة سنوات.

من هنا يمكن الاعتقاد بأن قيمة الصدق التنبؤي تنخفض عن قيمة الصدق التزامني، وتعد قيم معامل الصدق التزامني حدودا قصوى لمعامل الصدق التنبؤي. وهذا يعني أنه إذا كانت قيمة معامل الصدق التزامني (.70) مثلا، فإنه يتوقع أن يكون مقدار معامل الصدق التنبؤي أقل من ذلك إذا استخدم المحك نفسه في الحالتين.

المحك وخصائصه

ليس من الضروري أن يتشابه محتوى المحك مع محتوى المتنبىء فقد يمكن التنبؤ بنجاح الطلاب بدراستهم الجامعية (**محك**) من خلال أدائهم على اختبارات قبول معينة (**متنبىء**). أو نتنبأ بنجاح شخص في تسويق بضائع معينة (محك) باستخدام الاستجابات على استبيانات شخصية (متنبىء). وكلما كانت التنبؤات دقيقة كان الإختبار أكثر فائدة.

ونظرا للأهمية المترتبة على القرار المرتبط بالصدق المرتبط بالمحك، فإن صدق المحك وثباته أمران رئيسان في هذا السياق، بمعنى أن المحك يجب أن يكون صادقا وثابتا بدرجة كبيرة بصرف النظر عن محتواه. والمفروض كذلك أن يزودنا المحك بتقديرات تمتاز بالاستقرار مع مرور الزمن، وبأقل خطأ ممكن. كما أن استخدام أكثر من محك يقلل الخطأ بدرجة جوهرية. من ناحية أخرى، فإن قيمة معامل صدق الإختبار المرتبط بمحك (r_{xy}) لا تتعدى قيمة الجذر التربيعي لمعامل ثبات الإختبار المتنبىء ($r_{xx'}$). وبالرموز:

$$r_{xy} \leq \sqrt{r_{xx'}}$$

فمثلا إذا كانت قيم معامل الثبات للإختبار المتنبىء تساوي (.4) أو (.81) على التوالي، فإن قيم معامل الصدق المرتبط بمحك لا تتجاوز القيم (.63) و (.90) على الترتيب مع ثبات باقي المتغيرات.

ويعرف معامل الصدق الناتج بهذه الطريقة الصدق الذاتي Self-Validity ، وهذا ليس بديلا مقبولا لأي مؤشر صدق للإختبار، وتتراوح قيمته بين الصفر والجذر التربيعي لمعامل ثبات الإختبار المتنبىء (عودة، 2000).

كما أنه من المفروض أن يكون الأداء على المحك مستقلا عن أي معرفة مسبقة عن نوع المتنبيء أو حتى مستوى الأداء عليه، وذلك لتجنب ما يعرف بتلوث المحك Criterion Contamination . فإذا اطلع المشرف على البرنامج التدريبي على معلومات عن مستوى أداء الطلبة الملتحقين على الإختبار المتنبيء، فستكون هذه المعلومات ملوثة. وجدير بالذكر أن تلوث المحك ربما يزيد مقدار معامل الإرتباط بين المتنبيء والمحك، وبالتالي يؤدي إلى رفع قيمة معامل الصدق التنبؤي، ذلك عندما يلجأ المدربون إلى التركيز بشكل واضح على الملتحقين في البرنامج من المتفوقين فقط. الأمر الذي يؤدي إلى زيادة التباين في درجات الملتحقين جميعا، وهذا يؤدي إلى زيادة معامل الإرتباط ويحسن معامل الصدق بالنهاية.

من ناحية أخرى، قد يعمل تلوث المحك على تقليل معامل الإرتباط بين المتنبيء والمحك عندما يبدأ المدربون بالتركيز على الملتحقين ضعاف مستوى الأداء على المتنبيء، الأمر الذي يؤدي إلى رفع درجاتهم على المحك، وهذا يقلل من مقدار التباين في درجات الطلبة مجتمعين على المحك، وهذا يؤدي في النهاية إلى تقليل معامل الصدق.

وجدير بالذكر أنه لا بد أن يكون هناك ارتباط منطقي بين المحك من ناحية، والسلوك المتنبأ به من ناحية أخرى. فإذا كان المحك يشير إلى السلوك المتعلق بالقدرة الكمية لدى مجموعة من الطلبة مثلا ، فإنه من غير المنطقي أن يكون السلوك المتضمن في المتنبيء يتطلب قدرة على السباحة مثلا وذلك لضعف الإرتباط المتوقع بين القدرة الكمية والقدرة على السباحة.

أنواع المحكات

تم تصنيف المحكات حسب الوقت الذي يتم قياسها فيه في ثلاثة أنواع (Crocker and Algina, 1986):

(1) محكات فورية Immediate Criteria وهـــي محكـــات متـــوفرة وسهلة القياس كما في الدرجات التي تخصص للوقت المطلوب من سكرتيرة لطباعة ثلاثة رسائل رسمية بدقـــة، وهـــذا النـــوع مـــن المحكات لا يعد فعالا بدرجة كافية لاستخدامها في أغراض التنبؤ.

(2) محكات قصوى (ذروية) Ultimate Criteria وهـــي المحكـــات الأكثر أهمية على الرغم من صعوبة قياســـها أو حتـــى تعريفهـــا إجرائيا. وخير مثال على هذا النوع فعاليـــة التـــدريس teaching effectiveness، والكفايـــــة الجراحيـــــة للطبيـــــب surgical competence، وهذه المحكات بمثابة سمات ومفاهيم نظرية تحتاج إلى تعريف إجرائي ليسهل ملاحظتها وقياسها. فإذا كنا نتحدث عن محك أقصى مثل فعالية التدريس في برنامج إعداد معلمـــين، فـــإن قياسه يتم من خلال تكرار ملاحظة أداء الأشخاص الذين تخرجـــوا من البرنامج على مدى فترة زمنية طويلة نسبيا (3-5 ســـنوات). ولأن هذا محك غير عملي ويصعب قياسه، فإن المفروض استبداله بمحك آخر يعرف بـــالمحك الوسيطي:

(3) محكات الوسيطية Intermediate Criteria كما فـــي تقـــديرات المشرفين على عملية التدريس أو الاستشاريين في مجال الجراحـــة أثناء مرحلة التدريب العملي practicum في برنامج شامل لإعداد المعلمين أو تصنيف الجراحين.

تقدير مستوى الأداء على المحك من خلال الأداء على المتنبىء

يمكن تحليل البيانات المتعلقة بالصدق المرتبط بمحك بأكثر من طريقة، وهذا يعتمد على طبيعة البيانات المتعلقة بكل من المتنبىء والمحك، كونهما متغيران متصلين أو متقطعين أو أن أحدهما متصلا والآخر وثابا (متقطعا) على النحو التالي:

1. **عندما يكون المتنبىء والمحك متغيرين متصلين والعلاقة المفترضة بينهما خطية**، يتم حساب معامل ارتباط بيرسون بينهما، وهو معامل الصدق المرتبط بالمحك. ويمكن تفسير قوة هذا المعامل من خلال تربيع مقدار قيمة هذا المعامل ويعرف هذا بمعامل التحديد، كما يمكن مقارنته بمعاملات صدق أخرى. فعندما يكون معامل الإرتباط بين المتنبىء والمحك (.70) فهذا يعني أن (.49) من التباين بين هذين المتغيرين مشترك. أي أننا إذا عرفنا مستوى أداء الأفراد على المتنبىء فإننا نستطيع التنبؤ بما مقداره 49 ٪ من التباين في أداء هؤلاء الأفراد على المحك. عند ذلك يمكن التنبؤ بعلامة الطالب على المحك من خلال علامته على المتنبىء من خلال التعويض في معادلة الانحدار الخطي البسيط، على اعتبار أن (y') هي العلامة التي يتم التنبؤ بها على المحك، و (x) هي علامة الشخص على المتنبىء، و(r_{xy}) معامل الصدق التنبؤي (معامل الإرتباط بين المتنبىء والمحك) كما يلي:

$$y' = r_{xy} \left(\frac{s_y}{s_x}\right)(x - \bar{x}) + \bar{y}$$

حيث أن (s_x) و (s_y) انحراف لمعياري للعلامات على المتنبىء والمحك على الترتيب. و يمثل الرمز $\bar{x}$ المتوسط الحسابي للعلامـات على المتنبىء، بينما يمثل الرمز $\bar{y}$ المتوسط الحسابي للدرجات على المحك.

وقد اقترح ننالي (Nunnally, 1978) أن معامل الإرتبـاط (.40) بين المتنبىء والمحك يعد الحد الأدنـى لمعامـل الإرتبـاط المقبـول لأغراض التنبؤ الفعال. وتتوقف الدقة في التنبؤ علـى مقـدار الخطـأ المعياري في التنبؤ ($s_{y.x}$) كما في المعادلة التالية:

$$s_{y.x} = s_y \sqrt{1 - r_{xy}^{\ 2}}$$

حيث يشير s_y إلى الانحراف المعياري للدرجات على المحك. وتـزداد الدقة في التنبؤ كلما نقصت قيمة الخطأ المعياري للتنبؤ، وهذا يقل كلما ازداد مقدار معامل الإرتباط بين المتنبىء والمحك (معامـل الـصدق المرتبط بمحك).

2. لتحسين عملية التنبؤ على نحو أفضل يتم اللجوء إلـى زيـادة عـدد المحكات والمتنبئات بواقع (2–3). **وينبغي أن ترتبط تلك المتنبئـات ببعضها ارتباطا ضعيفا، وترتبط عاليا بالمحك، وبـذلك فـإن فـرص التنبؤ تتحسن ويقل الجهد المطلوب.**

3. **عندما يكون المتنبىء متغيرا متصلا كمعدلات الطلبة التراكمية نهاية المرحلة الجامعية ويكون المحك متغيرا ثنائيا كالنتيجـة فـي العمـل التدريسي (ناجح ، راسب)،** عندها يمكن تقدير الفرق بين متوسـطي درجات الطلبة على المتنبىء حسب تصنيفهم على المحـك باسـتخدام الأسلوب الاحصائي (t -test).

4. **عندما يكون المتنبىء والمحك متغيرين ثنائيين، كأن يكون المتنبىء بمثابة الأداء على اختبار قبول للترشيح للالتحاق في برنامج إرشادي (مقبول، غير مقبول) والقرار على المحك المتمثل بتقرير من مدير البرنامج عند الانتهاء منه (ناجح، راسب)**، فالأسلوب الإحصائي المستخدم للتعبير عن صدق المرتبط بالمحك هو معامل ارتباط فاي (phi- coefficient). وعندما يكون المتنبىء والمحك متغيرين متقطعين يتألف أحدهما أو كلاهما من مستويين أو أكثر فإن أسلوبا إحصائيا آخر يمكن استخدامه هو مربع كاي (chi-square). **وفي كلتا الحالين، فإن معامل الصدق المرتبط بمحك يسمى معامل دقة القرار Accuracy.**

		النتيجة على المحك	
		ناجح	راسب
النتيجة على المتنبىء	مقبول	a: قرار صائب	b: قرار خاطىء
	غير مقبول	c: قرار خاطىء	d: قرار صائب

ويعرف معامل دقة القرار بنسبة القرارات الصائبة، وهو حاصل قسمة (d+a) على (d+c+b+a) وجدير بالذكر أن (a) و (d) تمثلان القرارين الصائبين، بينما تمثل الخلايا (c) و (b) القرارين غير الصائبين كما في المعادلة التالية:

$$Accuracy = \frac{a+d}{a+b+c+d}$$

مثال: تقدم (100) شخص للمنافسة في الاشتراك ببرنامج تدريبي في الإرشاد النفسي، قبل منهم (80) شخصا لأن درجاتهم كانت على المتنبىء أعلى من الحد الأدنى المطلوب (علامة القطع)، وتم قبول الآخرين قبولا مشروطا . وبعد نهاية البرنامج نجح على الإختبار النهائي للبرنامج

(المحك) من الذين قبلوا (70) ولم يوفق منهم (10). ولم ينجح سوى (5) من الذين قبلوا قبولا مشروطا، أحسب دقة القرار الذي صدر بشأن عملية القبول (معامل الصدق):

		النتيجة على المحك	
		ناجح	راسب
النتيجة على المتنبىء	قبول كامل (80)	70	10
	قبول بشروط (20)	5	15

$$Accuracy = \frac{a+d}{a+b+c+d} = \frac{70+15}{10+70+15+5} = 0.85$$

وفي حالة أن تعليمات البرنامج لا تسمح بقبول الطلبة قبولا مشروطا، فإن معامل دقة القرار ستكون تبعا لذلك:

1- Accuracy = 70 / 80 = 0.88

جدول التوقع Expectency Table

هو جدول يصف العلاقة بين الدرجات على المتنبىء والدرجات على المحك. وبعبارة أخرى يساعد جدول التوقع في تقدير احتمالات نجاح الأفراد على المحك عندما يحصلون على درجات معينة على الإختبار المتنبىء، أو هو طريقة لتفسير صدق التنبؤ (Nitko, 1983).

يتم بناء مثل هذا الجدول بتحديد فئات الدرجات على الإختبار التنبؤي على المحور الأفقي، وتحديد أقسام الدرجات على المحك على المحور العمودي. ثم تدون النسب المئوية من الأفراد الذين ينتمون إلى فئة معينة

من فئات درجات المتنبىء وينتمون في الوقت نفسه إلى قسم معين من أقسام درجات اختبار المحك.

ويمثل الجدول التالي (تم اقتباسه من: Nitko, 1983) توزيعا لأعداد ونسب الطلبة المتقدمين لامتحان قبول للالتحاق ببرنامج تربوي معين، وقد صححت درجات المتقدمين على الإختبار المتنبىء من (100). بينما كان المعدل التراكمي للطلبة نهاية السنة الأولى من البرنامج يمثل المحك. وقد توزعت الدرجات عليه على النحو التالي: A, B, C, D, F

أ- التوزيع التكراري للدرجات على المحك حسب الفئات على المتنبىء

الفئات على المتنبىء	عدد الطلبة على كل درجة من درجات المحك					
	F	D	C	B	A	المجموع
80-89			1	3	1	5
70-79		1	4	5	2	12
60-69		3	5	6	1	15
50-59		4	8	5	1	18
40-49	1	5	8	4		18
30-39	1	6	5	3		15
20-29	2	5	4	1		12
10-19	1	2	1			4
0-9	1					1
المجموع	6	26	36	27	5	100

ب- جدول التوقع تم بتحويل أعداد الطلبة إلى نسب مئوية

الفئات على المتنبىء	نسب الطلبة على درجات المحك					
	F	D	C	B	A	المجموع

80-89			20	60	20	100
70-79		8	33	42	17	100
60-69		20	33	40	7	100
50-59		22	44	28	6	100
40-49	6	28	44	22		100
30-39	7	40	33	20		100
20-29	17	42	33	8		100
10-19	25	50	25			100
0-9	100					100

ويمكن قراءة هذا الجدول بالقول بأن 15 طالبا مثلا كانت درجاتهم تقع في الفئة بين 60 و 69 ، حصل ثلاثة منهم (بواقع 20 %) على الدرجة D ، بينما حصل خمسة منهم (بواقع 33%) على الدرجة C ، وحصل أربعون بالمائة منهم (6 طلاب) على الدرجة B ، وكان طالبا واحدا فقط (بواقع 7 %) قد حصل على الدرجة A.

وإذا تم قبول جميع الطلبة الذين حصلوا على درجة القطع Cut-off Score وهي هنا 60 فأكثر على اختبار القبول (المتنبىء)، وكانت درجة النجاح على المحك هي (B)، فيمكن حساب نسبة النجاح Success Ratio على النحو التالي:

عدد الطلبة الذين كانت درجاتهم على المتنبىء 60 وأكثر=

32=5 +12+15

عدد الطلبة الذين كانت درجاتهم على المحك B وأفضل=

18=1+2+1+3 +5+6

نسبة النجاح = 18 / 32 = 56 %

أما نسبة القبول Admission Ratio فهي حاصل قسمة عدد الذين كانت درجاتهم على المتنبىء تساوي وأكبر من درجة القطع (60) على عدد الطلبة الكلي (100). وبذلك تكون نسبة القبول= 32 %.

وينبغي هنا ملاحظة أن نسبة القبول تنخفض كلما زاد عدد المقبولين، كما ينبغي ملاحظة أنه كلما كان معامل الصدق تاما أي مساويا الواحد الصحيح، فإن جميع القيم في خلايا الجدول تتراكم بحيث تقع على خط مستقيم، وعندئذ يمكن التنبؤ بدرجة المحك بمعرفة درجة اختبار القبول تنبؤا تاما بغض النظر عن موقع درجة القطع. بذلك تكون نسبة النجاح واحدا صحيحا، أي لا يحتمل رسوب أي طالب تم قبوله. أما من رفض قبوله فيتوقع أن يرسب إذا تم قبوله عن طريق الخطأ (علام، 2000).

والجدير بالذكر فإن موقع درجة القطع يتحدد بناءا على حاجة العمل من الأفراد العاملين، وعدد الأماكن الشاغرة، ومدى الحاجة الملحة لشغلها. فإذا اتجهت النية لدى أجهزة الدولة ضرورة تخريج مجموعة من الموظفين لسد حاجات ماسة في دوائر معينة ، فإنه يصار إلى تخفيض قيمة درجة القطع، وبعكس ذلك، أو عندما تتجه النية في تلك الدوائر إلى تعيين أصحاب الكفاءات المميزة، فقد يصار إلى رفع قيمة تلك الدرجة.

يمكن تكوين جداول توقع لإختبارين من إختبارات التنبؤ بدلا من اختبار واحد، وهذا يقلل من الأخطاء التنبؤية للمهن والوظائف وغيرها. فمثلا، قد يتم استخدام اختبارات القبول بالإضافة إلى درجات نهاية العام في الثانوية العامة للتنبؤ بالمعدل التراكمي للطالب نهاية السنة الجامعية الأولى.

وبالرغم من شيوع استخدام جداول التوقع في عمليات الإختيار والتنبؤ، إلا أنه يعاب عليها استخدام فئات درجات اختبار التنبؤ لكل قسم من أقسام درجات المحك، مما يقلل الدقة في التفسير. إضافة إلى أن بيانات هذه الجداول تستمد من عينات تم اختبار أفرادها في الماضي، وربما تختلف خبراتهم عن خبرات الآخرين بعد ذلك. كما وأن تفسير القيم ينطبق على المجموعة ككل، ولا يتعلق بكل فرد على حدة، وبالتالي يتم التفسير بناءا على متوسط المجموعة التي ينتمي إليها الفرد (Allen and Yen, 1979).

الصدق المرتبط بمحك وحجم العينة

إن استخدام عينات صغيرة في تقدير صدق المحك التنبؤي أو التزامني يؤدي إلى زيادة الخطأ العيني. فقد أشار شميدت وهنتر وأري (Schmidt, Hunter, and Urry, 1976) إلى أن استخدام عينات بحجم يتراوح بين 30 و 50 يؤدي إلى أن يكون المتنبىء على مقدرة على التنبؤ بنسبة تتراوح بين 25 % و 35%. ومن ناحية أخرى، فإن عينة ممثلة وبواقع 200 فردا أو أكثر تحقق الوصول إلى صدق تنبؤي فعال بنسبة 90 % (بافتراض ضبط التجانس والمتغيرات الأخرى).

الصدق المرتبط بمحك والمدة الفاصلة بين تطبيق المتنبىء وتطبيق المحك

يفترض أن تكون الفترة الزمنية بين تطبيق المتنبىء وتطبيق المحك عملية ومناسبة، بحيث لا تكون قصيرة جدا لا تسمح بالحكم الكافي على فعالية المحك الذي يتوقع أن يساعد في اتخاذ قرار حول التنبؤ، وعندها يصبح معامل الصدق المرتبط بالمحك صدقا تزامنيا فقط.

من ناحية أخرى، فإن إطالة الفاصل الزمني كثيرا قد يؤدي إلى ازدياد احتمالية تأثير الأخطاء العشوائية في المحك. ويمكن تحديد الفترة الزمنية في ضوء التجارب البحثية، وكذلك بعد مراجعة الأدب البحثي ذي العلاقة، إضافة إلى الخبرات السابقة للخبراء والمتخصصين في هذا المجال (Crocker and Algina, 1986).

الصدق المرتبط بمحك وتجانس أفراد عينة الصدق

يفترض أن تكون عينة دراسة الصدق المرتبط بالمحك مراعية إلى درجة كبيرة خصائص الفئات التي سيتم تطبيق الدراسة عليها لأغراض التنبؤ مؤخرا. فقد يخشى أحيانا أن يصار إلى تطبيق دراسة الصدق على عينة من أفراد متجانسين فيزيد احتمال أن يكون معامل الصدق (الذي هو بمثابة معامل ارتباط) منخفضا. وهذا يشير إلى عدم الثقة بالقرار المعتمد

على نتائج هذه الدراسة. الأمر الذي بات من الضروري أن يتم اختيار عينة الصدق بعناية أكبر، لتحقق تمثيلا أفضل للمجتمع المعني بعملية التنبؤ، بمعنى أن تتمتع مجموعة الصدق بتباين معقول بين أفرادها، ولا تقتصر الدراسة على المقبولين فقط في البرنامج الخاضع لعملية التنبؤ بناءا على درجة القطع. هذا على الرغم من أن الإختبار المتنبىء ربما يتمتع بخصائص سيكومترية جيدة نسبيا (Mehrens and Lehmann, 1991).

الصدق المرتبط بمحك وعدد فقرات الإختبار المتنبىء

عرفنا أن مقدار معامل الصدق المرتبط بمحك يعتمد بدرجة كبيرة على معامل ثبات المحك، وأنه لا يتجاوز الجذر التربيعي لمعامل ثبات المحك. كذلك فإن نوعية وخصائص الإختبار المتنبىء يؤثر كذلك في مستوى الصدق، ومنها الثبات الذي يعتمد كثيرا على عدد الفقرات في الإختبار. فالإختبار الأطول نظريا هو الأقدر على تغطية جوانب المجال السلوكي موضع الاهتمام في عملية التنبؤ. وبشكل أدق، ليس المقصود هنا مجرد إضافة فقرات للإختبار لزيادة معامل الثبات وتحسين معامل الصدق، بل إن العملية تتعلق بإضافة فقرات ذات خصائص سيكومترية فعالة تزيد من مقدار التباين في الدرجات، وتعظم معامل الثبات الذي يعمل على تحسين مؤشرات الصدق تبعا لذلك (Brown, 1976).

كيف يتأثر معامل الصدق المرتبط بالمحك بمعاملي ثبات المتنبىء والمحك؟

عرفنا سابقا علاقة الصدق بالثبات من حيث العلاقة التكاملية بينهما من حيث أن أحدهما (الصدق) شرط لازم لحدوث الآخر (الثبات) وليس العكس يكون صحيحا بالضرورة. وبشكل أدق، فإن الزيادة في معامل الصدق مرهون بزيادة التباين الحقيقي العائد إلى السمة، والذي بدوره يؤدي إلى زيادة معينة في معامل الثبات. بينما عرفنا بأن الزيادة في معامل الثبات، قد تنسب إلى الزيادة في التباين الحقيقي العائد إلى السمة أو التباين

الحقيقي غير العائد إلى السمة. الأمر الذي ربما يكون سببا في زيادة معامل الصدق أو غير ذلك.

من جانب آخر، عرفنا بأن معامل الصدق المرتبط بمحك (التنبؤي أو التزامني) ما هو إلا معامل ارتباط بين المتنبىء والمحك (r_{xy}). وحتى يكون معامل الصدق ذا معنى ودلالة، لا بد من أن يكون كل من معامل ثبات المتنبىء ($r_{xx'}$) ومعامل ثبات المحك ($r_{yy'}$) عاليا ومقبولا بدرجة معينة. بمعنى أنه لا بد من توفر افتراض غياب الأخطاء العشوائية في كل من المحك والمتنبىء.

هذا الافتراض يجعل قيمة معامل الصدق تساوي (1)، وهذا طموح صعب التحقق. وما دام الواقع كذلك، فإن انخفاضا على مستوى معامل الصدق المحسوب سوف يحدث. والعلاقة التالية تبين العلاقة بين المتنبىء والمحك وثبات كل منهما:

$$r_{xy} \leq \sqrt{r_{xx'}}\sqrt{r_{yy'}}$$

وهذا يشير إلى أن معامل الصدق يتجاوز حاصل ضرب الجذر التربيعي لمعامل ثبات كل من المتنبىء والمحك (Lord and Novick, 1968; Ebel, 1968; Kane, 1982) . فمثلا يتوقع أن تكون أقصى قيمة لمعامل الصدق (.72) عندما يكون معامل ثبات المتنبىء (.81) والمحك (.64). **وربما يشذ عن هذه القاعدة عندما يسقط الفرض القائل بعدم ارتباط أخطاء القياس في الإختبارين.**

صدق البناء Construct Validity

تم تطوير فكرة صدق البناء حديثا مقارنة مع أنواع الصدق الأخرى (Cronbach and Meehl, 1955). وقد أمكن تعريف صدق البناء للإختبار بأنه الدرجة التي يعمل الإختبار على قياس خاصية أو سمة صمم أساسا لقياسها. والمعروف أن صدق البناء كباقي أنواع الصدق هو مجمل

إجراءات مستمرة ongoing ولا تنتهي. وتضع لوفنجر (Loevinger, 1979) صدق البناء في إطارين هما:

- إلى أي حد يقيس الإختبار خاصية أو سمة لها وجود فعلي ؟
- إلى أي حد يكون هناك تناظر بين التفسير المقترح للسمة أو للخاصية وما يقيسه الإختبار فعلا؟

ويتصل الإطار الأول بصدق الإختبار كما تبين سابقا، في حين يعرف الثاني بصدق التفسير. ولا شك أن التساؤلين يكمل بعضهما الآخر، إذ يقدم الإطار الأول أدلة حول مقدار ونوع الإرتباطات بين فقرات الإختبار، وهذا يضفي معنى سيكومتري على الإختبار. في حين يزود الإطار الثاني معنى عملي أو مهني (سيكولوجيا كان أو تربويا) للإختبار.

وإذا كان صدق المحتوى الأكثر مناسبة للإختبارات التحصيلية التي تمثل محتوى سلوكي محدد ويعالج سمة موجودة فعلا، وكان الصدق المرتبط بالمحك يتعلق بدقة القرار المتعلق بقدرة الإختبار على التنبؤ بمستوى أداء أو سلوك معين. فإن صدق البناء يرتبط ببناء أدوات تتحقق من وجود قدرة عقلية أو سمة نفسية من ناحية، وكذلك قياسها بدقة من ناحية أخرى. وإذا كان الصدق المرتبط بالمحك يخدم القرار التنبؤي بصرف النظر عن محتوى أداة القياس (المتنبىء)، فإن صدق البناء كصدق المحتوى يهدف إلى خدمة أداة القياس نفسها.

خطوات التحقق من صدق البناء

تركزت عملية جمع الأدلة حول صدق البناء للإختبار حول عملية بناء نظرية مصغرة miniature تتعلق بالإختبار من خلال الخطوات التالية (Crocker and Algina, 1986; Cronbach and Meehl, 1955):

1. تبني نظرية معينة يستند إليها الإختبار ويشتق منها فرضية أو فروض تبين كيف يختلف أو يتشابه الأفراد في امتلاكهم للخاصية

أو السمة في ضوء ما يحصلون عليه من درجات مختلفة على الإختبار.

2. اختيار أو تطوير أداة قياس تتألف من فقرات تمثل سلوكات محددة تصف السمة
3. جمع بيانات فعلية ميدانية Empirical Data بقصد إختبار الفرضيات.
4. تقييم ما إذا كان هناك انسجام بين البيانات الفعلية وإختبار الفرضيات، وما إذا كان تفسير المواقف والنتائج يتم بواسطة النظرية. وبغير ذلك، يجب مراجعة النظرية أو المقياس أو تكرر الخطوات السابقة للتوصل إلى تفسير أكثر فائدة ودقة، ولاختبار فرضيات أخرى.

وبناءا على النظرية التي انطلقنا منها والتي يعتقد بأنها تفسر أو يستند إليها في تطوير المقياس السمة، فإن مطورو الإختبار يؤسسون تنبؤات حول كيفية استخدام نتائج الإختبار في أوضاع ومواقف مختلفة. وهذه التنبؤات ستخضع للفحص المنطقي والإحصائي. فإذا كانت البيانات تدعم تلك التنبؤات حسب نتائج فحص الفرضيات، فإن الإختبار صادق البناء. وغير ذلك، فستبرز ثلاثة احتمالات:

(1) إما أن التجربة تعاني من ضعف أو خلل عندما صممت لجمع البيانات، أو

(2) أن هناك خطأ أو قصورا في النظرية ويجب دراستها أو تعديلها تبعا لذلك، أو

(3) أن الإختبار لا يقيس السمة موضع البحث.

أساليب التحقق من صدق البناء

يمكن تلخيص الخطوات والعمليات التي تشكل بمجموعها أساليب التحقق من صدق البناء لاختبار يقيس خاصية معينة على النحو التالي:

- **التحليل المنطقي Judgmental Analysis**: فقد يمكن دراسة صدق البناء لاختبار يقيس سمة عقلية أو نفسية من خلال الفحص الدقيق لفقرات وتعليمات الأداة وكذلك تحليل نوع الأداء المطلوب وتفسيره. كما يتوقع أن يتم إحداث تكامل بين نتائج الفحص والنظرية التي يستند إليها الإختبار وآراء المختبرين الذين سبق أن اختبروا باختبارات مشابهة. ويترتب على ذلك أحيانا افتراض تفسيرات ربما تختلف عن التفسيرات التي اقترحها معد الإختبار أساسا.

  وختاما، يعد التحليل المنطقي مصدرا مهما من مصادر التوصل إلى فروض بديلة تتعلق بالأداء على الإختبارات والمقاييس، فالمحكم الذي لديه خبرة سابقة ومتعمقة بالأخطاء التي شابت الإختبارات السابقة يمكنه أن يكتشف جوانب الضعف في أداة القياس الجديدة.

- **الفروق بين المجموعات Group Differences**: فقد تتضمن النظرية وجود أو عدم وجود فروق في الدرجات على الإختبار، الأمر الذي يستوجب عندها اختبار ذلك إحصائيا بناءا على البيانات التي يتم جمعها من الميدان. فقد يتم اختبار الفروق بين الجنسين (ذكور وإناث) أو الفروق بين الأفراد (مراهقين و أطفال) أو الفروق بين درجات الطلبة حسب ثقافات الممتحنين على اختبار قدرة عقلية (الذكاء) أو سمة نفسية (الاكتئاب)، وفيما يلي نماذج لفرضيات صفرية حول هذا السياق (بالرغم من أن الفرضيات في صدق البناء ينبغي أن تصاغ بصورة متجهة:

  - لا يختلف مستوى ذكاء طلبة الصف العاشر باختلاف جنسهم.

- لا يختلف مستوى ذكاء الأطفال حسب خلفياتهم الاجتماعية والاقتصادية.
- لا يختلف مستوى ذكاء الأطفال باختلاف الفئة العمرية.

- **التغيرات** Changes: فقد تتضمن النظرية وجود تغير ما في درجات الممتحنين مع التغير في الزمن (العمر). فربما تفترض النظرية أن مستوى الذكاء أو مهارات التواصل الشفهي لدى الأطفال تتغير (تنمو) مع العمر. ولا بد من فحص هذا التغير، والصياغة التالية نموذج لفرضية متجهة تختبر مدى التغير الجوهري في مستوى الذكاء على مدى فئات عمرية متتابعة (6–16) سنة مثلا:

ينمو مستوى أداء الأطفال على مقياس الذكاء جوهريا مع تقدمهم في العمر بين سن 6 و سن 16 سنة على مستوى الدلالة الاحصائية (.05)

- **الإرتباطات** Correlations: فقد تنتهي النظرية بمعاملات ارتباط موجبة، أو سالبة، أو حتى صفرية بين السمة التي يقيسها الإختبار وعدد آخر من السمات. فقد ترتبط الدرجات على اختبار الذاكرة القصيرة ارتباطا موجبا مع العمر ولا تختلف (ترتبط) باختلاف الجنس. ولا يقتصر ذلك على معرفة العلاقة الموجبة بين الإختبار الذي يجري تطويره وغيره من الإختبارات المشابهة، بل المفروض التعرف على علاقة ذلك الإختبار ببعض الإختبارات والمقاييس التي لا تقيس السمة. وبهذا الصدد اقترح كامبل وفيسك (Campbell and Fiske, 1959) أسلوبا للتحقق من ذلك اعتمادا على مصفوفة أطلقا عليها مصفوفة السمات المتعددة والطرق المتعددة Multitrait - Multimethod Matrix يمكن استخدامها في دراسة صدق البناء في اتجاهين هما:

- **الصدق التقاربي Convergent Validity** للمقياس عندما تنتج معاملات ارتباط موجبة وعالية بين الدرجات على مقياس السمة موضع البحث والدرجات على اختبارات تقيس سمات مشابهة.
- **الصدق التباعدي Divergent Validity** للمقياس عندما يكون معامل الإرتباط بين الدرجات على السمة التي يقيسها المقياس موضع الإهتمام والدرجات على اختبارات مختلفة عنه ضعيفا أو سالبا.

فلو فرضنا أن هناك سمتين (س$_1$ ، س$_2$) يتم قياس كل منها بطريقتين (ط$_1$: الاختيار من متعدد، ط$_2$: الاختيار من بديلين (صح أو خطأ)) ، فهذا يعني أن هناك أربعة اختبارات تطبق على عينة من الأفراد. ويتوقع الحصول على مصفوفة من معاملات الإرتباط كما في الشكل التالي (تم اقتباس القيم في خلايا المصفوفة من (Allenand Yen, 1979):

		ط$_1$		ط$_2$	
		س$_1$	س$_2$	س$_1$	س$_2$
ط$_1$	س$_1$	.80	.20	.75	.17
	س$_2$		.85	.15	.71
ط$_2$	س$_1$			.83	.21
	س$_2$				.91

فمعاملات الإرتباط الناتجة من قياس السمة الواحدة بطريقتين مختلفتين من مثل (.75) و (.71) تعرف بمعاملات **الصدق التقاربي**. بينما المعاملات الإرتباطية الناتجة من قياس سمتين بطريقة واحدة أو قياس سمتين مختلفتين بطريقتين مختلفتين تعرف بمعاملات **الصدق التمايزي** كما يلي:

- السمة (س$_1$) والسمة (س$_2$) باستخدام الطريقة ط$_1$ = .20

- السمة (س$_1$) والسمة (س$_2$) باستخدام الطريقة ط$_2$ = 21.
- السمة (س$_1$) بطريقة ط$_1$ والسمة (س$_2$) بطريقة ط$_2$ = 17.
- السمة (س$_1$) بالطريقة ط$_2$ والسمة (س$_2$) بالطريقة ط$_1$ = 15.

وجدير بالذكر أن المعاملات التي تقع في قطر المصفوفة من مثل (.80) و (.85) تمثل معاملي ثبات السمة الأولى والسمة الثانية على الترتيب باستخدام الطريقة الأولى، بينما يمثل المعاملان (.83) ، (.91) معاملي ثبات السمة الأولى والثانية على الترتيب باستخدام الطريقة الثانية. كما يلاحظ من قراءة هذا الجدول أن قيم معاملات الصدق التقاربي أكبر من قيم معاملات الصدق التمايزي.

- **التجريب Experimenting** : يمكن دراسة صدق الإختبار عن طريق إحداث تغيرات في درجات الأفراد في اختبار ما كوسيلة للتعرف على ما إذا كان يتأثر الأداء بمعالجات أو متغيرات معينة، وهذا ربما يعمل على تأكيد بعض التفسيرات المتعلقة بنتائج الإختبار أو رفضها. فقد يمكن التحقق من صدق البناء لمقياس "القلق" مثلا، عن طريق تحديد ما إذا كان المقياس يزودنا بنتائج تتفق مع نظرية معينة تتفق مع نظرية معينة تتعلق بالقلق. فإذا أوضحت النظرية أن القلق يظهر بوضوح في مواقف معينة تثير الإحباط فإنه يمكن إجراء تجربة لتحديد ما إذا كان الأفراد الذين يواجهون أحد هذه المواقف يحصلون على درجات مرتفعة في المقياس أكثر من الذين لم يواجهوا الموقف.

ومن ناحية أخرى، إذا أشارت النظرية إلى أن سمة ما لا تتأثر بالنضج أو المتغيرات البيئية فإنه يمكن التحقق من ذلك تجريبيا. ففي دراسة صدق البناء لاختبار يقيس سمة الذكاء العام ، وكنا نفترض أيضا بأن الذكاء يزداد بسرعة لدى الأطفال، بينما ينمو بمعدل أقل لدى المراهقين، ويقل ازدياده أكثر حين بلوغ سن النضج. فإننا نطبق اختبار الذكاء موضع البحث

على عينات من الأفراد من سن (5) أعوام إلى (30) عاما. ومن ثم يصار إلى رسم منحى يمثل متوسطات درجات كل مجموعة عمرية. فإذا أكد المنحنى الفرض السابق، فهذا دليل على صدق البناء للإختبار (علام، 2000).

الصدق العاملي Factorial Validity

وهذا شكل من أشكال صدق البناء يتم الوصول إليه من خلال التحليل العاملي Factor Analysis. والتحليل العاملي أسلوب رياضي يمثل عددا كبيرا من العمليات والمعالجات الرياضية في تحليل الإرتباطات بين المتغيرات (فقرات المقياس أو الإختبار) ومن ثم تفسير هذه الإرتباطات واختزالها في عدد أقل من المتغيرات تدعى عوامل Factors.

والجدير بالذكر أن التحليل العاملي كمنهج رياضي في التحليل يعد أسلوبا ليس من السهولة تناوله وفهمه، بل هو أسلوب معقد في الأصل، ساعد التقدم في برامج الحاسوب في انتشاره وسهولة استخدامه في مجالات متعددة وخاصة في تطوير المقاييس والإختبارات بأنواعها (Allen and Yen, 1979). وبشكل محدد، يساعد التحليل العاملي في الحكم على أن السمة التي يقيسها الإختبار ذات بعد واحد أو متعددة الأبعاد. ففي بداية التحليل يمكن التوصل إلى أن السمة المقاسة في الإختبار أحادية أو متعددة من خلال مطالعة قيم الجذر الكامن Eigen Value للعوامل الناتجة، إذ أن القيمة الأدنى للجذر الكامن لأي عامل على درجة من الأهمية هي (1) أو أكثر (حسب معيار كايزر). عند هذه الخطوة يمكن الوصول إلى أن السمة أحادية (إذا كان هناك عامل واحد فقط) أو أنها ثنائية البعد أو أكثر إذا كان هناك عاملين أو أكثر، قيمة الجذر لكل منهما (1) أو أكثر كما في الجدول التالي:

Factor	eigen value	percent of variation
1	5.7065	40.8 %
2	2.3554	16.8 %
3	2.0095	6.4 %
4	0.8897	5.1 %
5	0.6897	3.9 %

ومن قراءة هذه البيانات، يمكن الاستنتاج أن المقياس يتألف من ثلاث عوامل قيم الجذر الكامن لكل منها أكثر من الوحـدة (5.700 ، 2.350 ، 2.0095). وجدير بالذكر أن يظهر عدد غير قليل من العوامـل كنتيجـة للتحليل العاملي الأولي، كما أن أهمية العامل تتحدد من خلال نسبة التباين الذي يشرحه العامل الواحد. ويتضح من هذه البيانات أن العوامل الثلاثـة تشكل 64 % من مجمل التباين.

ولتحديد ما إذا كان التحليل العاملي يفضي إلي تركيب أحادي العامل أو متعددها، فإنه إذا كانت نسبة الجذر الكامن للعامل الأول تفوق ضـعف تلك النسبة للعامل الثاني، أو أن نسبة الجذر الكامن للعامل الأول إلي العامل الثاني تفوق مثلي نسبة العامل الثاني إلي الثالث. فيمكن القول بأن التحليـل يفيد بوجود اختبار أحادي البعد.

والخطوة التالية تتلخص في تصنيف فقرات المقياس وتوزيعها علـى العوامل الناتجة. إذ أن هناك معامل ارتباط بين كل فقرة من فقرات المقياس وكل عامل من العوامل الناتجة، وتعرف هذه المعـاملات **بتـشبع الفقـرة بالعامل factor loading** ، يتبع هذه المرحلة عملية تدوير لتلـك العوامـل لتظهر تشبع الفقرات على العوامل الناتجة بشكل يخضع لقاعدة البناء البسيط والتي تضمن تفرد الفقرات في تمثيل العوامل ، لأن المهم في بناء المقياس أن يقل احتمال مساهمة الفقرة في قياس أكثر من عامل (مجال). والجـدول التالي يبين تشبع عينة من سبع فقرات variables على ثلاثة عوامل:

Variable	Factor 1	Factor2	Factor 3
v_1	-.164	-.622	-.337
v_2	.719	-.047	.233
v_3	.807	.279	-.145
v_4	.199	.357	-.407
v_5	-.518	.178	.694
v_6	-.274	.863	.229
v_7	-.003	.491	.330

وعادة ما يتم تدوير هذه العوامل بطريقة معينة يعاد توزيع تشبع الفقرات على العوامل المستخرجة بشكل يسهل تبلور عوامل ذات معنى. وهناك طرق متعددة للتدوير Rotation ، فهناك تدوير متعامد من مثل **تدوير فاريماكس Varimax** وتدوير مائل من مثل **Oblique** عندما تكون العوامل مترابطة وهكذا. والجدول التالي يبين نتائج تشبع الفقرات على العوامل بعد التدوير المتعامد من نوع فاريماكس:

Variable	Factor 1	Factor 2	Factor 3
v_1	-.135	.009	-.822
v_2	.866	-.142	.089
v_3	.782	.176	.230
v_4	-.058	.794	.151
v_5	-.397	-.754	.144
v_6	.174	.229	.842
v_7	-.121	-.006	.777

ويمكن من هذا الجدول الاستنتاج بأن الفقرات (1) و (6) و (7) تتشبع على العامل الثالث، بينما تتشبع الفقرات (4) و (5) على العامل الثاني، والفقرات (2) و (3) تتشبع على العامل الأول. كما يستنتج أن السمة التي يقيسها الإختبار تتكون من ثلاثة أبعاد فرعية.

يلي هذه العملية تفحص محتوى الفقرات التي تنتمي للعامل الواحد، ودراسة ما هو مشترك بين هذه الفقرات، ليصار لاحقا إلى تسميتها. كما سيجري مقارنة العوامل الناتجة مع العوامل التي تنص عليها النظريات التي تفسر وجود السمة موضع البحث، فقد يتم الاتفاق بين ما هو قائم، عندئذ يكون دور **التحليل العاملي تأكيديا** **Confirmatory Factor Analysis**، أو ربما يقوم بدور ريادي استكشافي لم يكن معروفا من قبل، وعندها يعرف التحليل العاملي **بالتحليل الاكتشافي** **Explanatory Analysis** (والجدير بالذكر أن التحليل العاملي الاستكشافي غير مرتبط في إجراءات صدق البناء، وإنما في البحوث والدراسات).

الصدق المتقاطع (الصدق عبر العينات) Cross Validation

يظهر في بعض الأحيان أن معامل صدق المحك التنبؤي ضعيفا إذا طبق على أفراد مجموعة معينة، على الرغم من احتواء المتنبىء على فقرات تميز جيدا بين أفراد مجموعة أخرى. وهذا يشير إلى وجود تذبذب في مقدار معامل الصدق من مجموعة إلى أخرى. الأمر الذي يتطلب تطبيق دراسة الصدق على عينة أو عينات أخرى من الأفراد. وفي حالة تشابه نتائج الدراسات فإن ذلك يزيد من موثوقية معامل الصدق، أما إذا اختلفت تلك النتائج ، فيجب العمل على تطوير فقرات المتنبىء وتحسين مواصفاتها (Crocker and Algina, 1986).

التداخل بين مؤشرات الصدق

على الرغم من تناول مؤشرات الصدق الثلاثة بشكل منفصل، فإنها عمليا يكمل بعضها الآخر. فقد يدعم الصدق التنبؤي إجراءات كل من صدق البناء وصدق المحتوى للإختبار. كما قد تستخدم نتائج الأداء على اختبار يقيس سمة معينة كاختبار متنبىء (Nunnally, 1978).

يظن البعض أن الصدق يرتبط بالصيغة النهائية للإختبار. هذا ليس دقيقا، إذ يتعلق صدق المحتوى مثلا بالمرحلة الأولى في بناء الإختبار، وتتمثل هذه المرحلة بتجميع الفقرات التي تمثل المجال السلوكي المنوي قياسه. ويمكن انتقاء عينة الفقرات باستخدام أساليب الصدق المرتبط بمحك من حيث تمييزها بين الناجحين والراسبين، أو بين المتمكنين وغير المتمكنين في ضوء محك ما. وبالتالي ، يمكن استبعاد الفقرات التي لا تحقق متطلبات سمة معينة تقيسها الفقرات، وربما يمكن التحقق من ذلك أيضا بقياس معاملات الاتساق الداخلي.

وجدير بالذكر أن ما ينطبق على اختبارات التحصيل ينطبق على اختبارات الذكاء والاستعدادات، وكذلك مقاييس الجوانب الوجدانية والشخصية. ولا يخفى أن الإجراءات المتعلقة بصدق البناء للسمة موضع البحث، يرتبط بالأدلة المتعلقة بصدق المحتوى لأداة القياس من ناحية، وصدقها المرتبط بالمحك من ناحية أخرى.

وفي الختام، فإن الصدق والثبات مفهومان مترابطان. إذ، حتى يكون المقياس ثابتا، فيجب أن يكون صادقا. حيث أن الزيادة في درجة الصدق تؤدي إلى زيادة حتمية في معامل الثبات، لكن العكس ليس شرطا أن يكون صحيحا.

صدق الإختبارات محكية المرجع

كانت مسألة الصدق- شأنها شأن قضية الثبات- في الإختبارات المحكية المرجع، مصدراً اختلافيا بين المتخصصين في العلوم السلوكية، والخبراء في القياس النفسي على وجه الخصوص. وطالما كانت الإختبارات المحكية المرجع تستخدم أساساً في مجال الإختبارات التحصيلية، يكون صدق محتوى الإختبار من أهم النقاط موضع البحث.

وفي هذا الخصوص، من المفروض أن يطابق صدق الإختبارات محكية المرجع درجة الصدق لدى الإختبارات معيارية المرجع أو يفوقها. وسبب ذلك يعود إلى أن العناصر السلوكية أو مفردات التحصيل تكون محددة أكثر في الإختبارات المحكية المرجع.

وبكل الأحوال، إذا أريد استخدام نتائج الإختبارات محكية المرجع في صنع القرارات التعليمية يجب التأكد من توفر مؤشرات عملية Empirical عن الصدق المرتبط بالمحك. لأننا دائماً "نرغب في التعميم إلى ما هو أكثر اتساعا من المجتمع الإحصائي المعرف عند بناء الفقرات" (Jackson, 1970).

وعلى سبيل المثال، إذا استخدمت نتائج اختبار إتقان أو تمكن Mastery Test لتحديد مدى إمكانية السماح لطالب ما في الانتقال إلى الوحدة الدراسية اللاحقة، يجب توفير بعض الأدلة لبيان أن الطلبة الذين كان تحصيلهم دون درجة التمكن Cutoff Score لم يظهروا تحسنا في الأداء في المادة اللاحقة، كما هو حال الطلبة الذين كان تحصيلهم يتجاوز تلك الدرجة.

إن استخدام الطرق الإرتباطية لتقدير درجة الصدق المرتبط بالمحك (التنبؤي أو التزامني) للإختبارات محكية المرجع، يجعلها ضعيفة، بسبب

ضعف التباين بين الدرجات على الإختبارات محكية المرجع (كما نوقشت سابقاً)، كما يتوقع أن تؤثر في معامل الثبات، وتجعله ضعيفا. لذلك، فإنه يصار إلى استخدام الطرق الإحصائية التمييزية عوضا عن الطرق الإرتباطية في حساب معاملات صدق الإختبارات محكية المرجع، وتصبح أكثر فائدة (Hambleton and Novick, 1973; Harris, et. al. 1974; Millman, 1974; Swaminathan et. al, 1975; Linn, 1980).

يقصد بصدق الإختبار محكي المرجع مدى ملاءمة الدرجات المستمدة من الإختبار للاستخدامات المناسبة للغرض الذي بني من أجله الإختبار. وهناك ثلاث طرق رئيسة لتقدير صدق الإختبار محكي المرجع تماثل إلى حد ما الطرق المتبعة لتقدير صدق الإختبارات معيارية المرجع وهي الصدق الوصفي، الصدق الوظيفي، وصدق انتقاء النطاق السلوكي. والتي تم تناولها تاليا:

الصدق الوصفي Descriptive Validity : وهذا يشبه صدق محتوى الإختبارات معيارية المرجع من حيث المفهوم وطريقة التحقق، بالرغم من أن مفهوم الصدق الوصفي أكثر شمولا من صدق المحتوى. وقد اقترح كرونباك Cronbach الاستعانة بمجموعتين من الخبراء الاختصاصيين في كتابة الفقرات، حيث يطلب من كل مجموعة على حدة كتابة فقرات في ضوء تحديد معالم النطاق السلوكي الذي يتوقع أن يمثله الإختبار، وكذلك المواصفات التفصيلية للإختبار التي تتضمن تحديدا للمحتوى، وقواعد فحص الفقرات ومراجعتها.

ولدى انتهاء المجموعتين من انجاز المهمة المطلوبة، يتم الوصول إلى اختبارين متوازيين أو متكافئتين. ويمكن التحقق من ذلك بتطبيق كل من الإختبارين على مجموعة واحدة من المفحوصين الذين تلقوا تعليما متعلقا بالمحتوى.

الصدق الوظيفي Functional Validity: ويرتبط ذلك بقدرة الإختبار على التنبؤ بنجاح الفرد في عمل معين. ويقصد بذلك أن يؤدي الإختبار الوظيفة التي بني من أجل تحقيقها. ويشبه هذا الصدق ما يعرف بالصدق التجريبي (الصدق المرتبك بالمحك) في حالة الإختبارات معيارية المرجع. بالرغم من أن تسمية الصدق بالوظيفي أكثر شمولية. إذ أن الصدق التجريبي يشير إلى قدرة الإختبار على التنبؤ بأداء الفرد في مواقف تختلف إلى حد ما عن المواقف التي يقيسها الإختبار. ونظرا لأن الإختبارات محكية المرجع لها وظائف واستخدامات متنوعة تهم المربين وبعضها لا يتضمن التنبؤ بمحكات معينة، فإن مفهوم الصدق الوظيفي ربما يكون أكثر انطباقا على هذه الإختبارات.

من ناحية أخرى، يختلف الصدق الوظيفي عن الصدق المرتبط بالمحك في أن الدرجات على الإختبار محكي المرجع تكون في العادة ثنائية (أفراد متمكنين، أفراد غير متمكنين). وعليه، تعكس مؤشرات الصدق الوظيفي مدى اتساق القرارات المستندة على الدرجات الثنائية لكل من الإختبار من جهة، والمحك الخارجي المستقل بدلا عن معاملات الإرتباط بينهما التي تستخدم في تقدير الصدق المرتبط بالمحك للإختبارات معيارية المرجع.

وبشكل عام، يجدر القول بأن الصدق الوصفي يعد شرطا لازما للصدق الوظيفي. ويلجأ البعض إلى التقليل من الصدق الوصفي للإختبار في سبيل زيادة الصدق الوظيفي للإختبار. أي أنه لكي يكون للإختبار فائدة قصوى في أغراض التقويم يمكن أن يشتمل الإختبار على أي مجموعة من الفقرات ما دام الإختبار يؤدي الغرض بفاعلية على أساس النتائج التجريبية. وهذا يتفق مع حالة الإختبارات معيارية المرجع التي تستخدم في أغراض التنبؤ بغض النظر عن التحديد الدقيق للنطاق السلوكي الذي يقيسه الإختبار.

لذلك بات من الممكن تقليل أو حتى إغفال أهمية الصدق الوصفي كشرط لازم للصدق الوظيفي للإختبارات محكية المرجع.

صدق انتقاء النطاق السلوكي Behavioral Domain Validity: وهذا يشبه إلى حد بعيد صدق البناء المستخدم في حالة الإختبارات معيارية المرجع. وتبرز أهمية هذا النوع من الصدق في أنه يرتبط بصدق تفسير درجات الإختبار، ومن ثم الاستدلال من تلك الدرجات على أداء الفرد في نطاق سلوكي أكثر اتساعا من عينة السلوك التي يقتصر الإختبار على قياسها.

يرتبط صدق انتقاء النطاق السلوكي بمفهوم إمكانية التعميم الذي نادى به كرونباك في نظرية التعميم Generalizability Theory التي تناولت هذا المفهوم. فالنطاق السلوكي الذي نختاره من بين النطاقات السلوكية الممكنة يجب أن يسمح لنا بالتعميم على النطاق الشامل للبعد الذي يقيسه الإختبار.

ولتوضيح ذلك، نفترض أننا أردنا قياس قدرة المتعلم على التعلم على تحليل الأفكار التي تنطوي عليها القطع النثرية تحليلا سليما. وقمنا بتصنيف السلوك الذي تتضمنه هذه القدرة إلى أربعة أقسام فهنا يجب أن ننتقي من بين هذه الأقسام الأربعة النطاق السلوكي الذي لو أتقنه المتعلم يمكن تعميمه على النطاقات الثلاثة الأخرى. ويمكن التحقق من ذلك بتجربة استطلاعية بعد بناء أربعة اختبارات محكية المرجع يقيس كل منها قسما من هذه الأقسام طبقا لمواصفات تفصيلية يتم إعدادها لكل اختبار. ويتم تطبيق الإختبارات الأربعة على عينة كبيرة من الطلاب الذين يوثق في قدرتهم على تحليل القطع النثرية. ثم يتم دراسة النسب المئوية للطلاب الذين حصلوا على الدرجات النهائية أو درجات مرتفعة في أي من الإختبارات

الأربعة، ويصار إلى مقارنة توزيع درجاتهم في الإختبارات الثلاثة الأخرى.

بذلك يتم تحديد النطاق السلوكي الذي يتميز بإمكانية تعميم أداء الطلاب فيه على النطاقات السلوكية الأخرى. وبالطبع يصعب إجراء هذه الطريقة عمليا لما تتطلبه من جهد ووقت كبيرين. لذلك اقترح بوبهام Popham طريقتين يمكن استخدام أي منهما للتوصل إلى أدلة عن صدق انتقاء النطاق السلوكي للإختبار وهما:

1. القيام بتدريس المعارف والمعلومات التي يتضمنها في كل نطاق من النطاقات السلوكية المطلوبة لمجموعات صغيرة من الطلاب حتى يصلوا إلى درجة التمكن، ثم اختبار إمكانية تعميم إتقان نطاق منها إلى النطاقات الأخرى.
2. الموازنة بين أداء مجموعتين من الطلاب في عدد من الإختبارات تقيس المعارف والمهارات التي يتضمنها كل نطاق من النطاقات السلوكية المطلوبة. ويجب أن تكون إحدى المجموعتين قد تلقت التعليم المتعلق بهذه النطاقات، والأخرى لم تتلق التعليم، ثم اختيار النطاق السلوكي الذي يميز بين هاتين المجموعتين.

والجدير بالذكر هنا، أن استخدام الطرق الإحصائية المعروفة في تقييم صدق بناء الإختبارات معيارية المرجع مثل التحليل العاملي، وتحليل السمات المتعددة والطرق المتعددة لا تناسب التوزيعات المتجانسة التي يكون مدى درجاتها ضيقا. بل يمكن الاعتماد على طرق أخرى مثل التحليل الهرمي لجوتمان Guttman Scalogram Analysis في تقييم صدق انتقاء النطاق السلوكي للإختبارات محكية المرجع.

ملخص

أمكن تلخيص الأفكار والمفاهيم التي تضمنها الفصل في النقاط التالية:

- الثبات ضروري للصدق، ولكنه ليس كافيا لبلوغه.
- يمكن تعريف الصدق على أنه مدى إمكانية صنع استنتاجات معينة بناءا على درجات الإختبار أو المقياس. وطالما يمكن استخدام الإختبار لأغراض مختلفة، فليس هناك معامل صدق واحد للإختبار.
- مؤشرا كل من الصدق والثبات نسبيان وموقفيان. فالإختبار يكون صادقا بالنسبة لأهداف معينة، وفي مجال معين، ولدى أفراد مجتمع معين، وليس بالضرورة أن يكون الإختبار صادقا وثابتا بشكل مطلق.
- يصنف الصدق في ثلاث هي: صدق المحتوى، الصدق المرتبط بالمحك، وصدق البناء.
- يرتبط صدق المحتوى بمدى دقة تمثيل محتوى الإختبار للمجال الذي يراد التوصل إلى استنتاجات بخصوصه.
- يعبر عن الصدق المرتبط بالمحك بأساليب دراسة العلاقة بين الدرجات على الإختبار وبين المحكات الخارجية المستقلة.
- من أجل دراسة الصدق المرتبط بالمحك، من المهم أن يكون لدينا مقياساً جيداً للمحك. ويجب أن يكون هذا المقياس مناسبا، وذا علاقة، وخالياً من الانحياز.
- صدق البناء هو مدى إمكانية تفسير الدرجات على الإختبار من خلال بعض البنى التفسيرية.

- الصدق المنهاجي والصدق التدريسي مصطلحان يستخدمان للتعبيـر عن مدى تغطية مادة الإختبار في المنهاج الدراسي، أو ما تم تدريسه فعلاً من بنود المنهاج. وهما ليسا من أنواع الصدق حسب المفهـوم التقليدي للمصطلح.
- هناك عدد من الطرق التي يعبر فيها عن الـصدق مثـل: معامـل الإرتباط، معامل التحديد، الخطأ المعياري فـي التقـدير، وجـداول التوقعات.
- تؤثر عوامل عديدة على الصدق، كما في معاملات ثبات كـل مـن المحكات والمتنبئات، ودرجة تجانس المجموعة، وشكل (أو نمـط) العلاقة بين المحك والمتنبىء.
- هناك عدة عوامل تؤثر في صدق الإختبار كونه يستخدم في صـنع القرارات، كتوفر بيانات متنوعـة، وكلفـة الإختبـار، والقـرارات الخاطئة، ونسبة الاختيار، ونسبة النجاح.
- توجد كمية لا بأس بها من الأدلة التي تفترض أن بإمكان الفرد، في الكثير من الحالات، أن يعمم من دراسة الصدق المرتبط بالمحك التي أجريت في موقف معين على مواقف أخرى.
- يجب أن يطابق صدق المحتوى في الإختبارات المحكية المرجـع أو يتفوق على صدق المحتوى في الإختبارات المعيارية المرجع.
- ليست الطرق الإرتباطية ملائمة لتحديد صدق الإختبـارات محكيـة المرجع، لأن تباين توزيع الدرجات يميل إلى الانكماش. لذلك، تكون الطرق الإحصائية التمييزية هي الأكثر استخداما.

أسئلة وتمارين

س1: أجب بـ (نعم) أو بـ (لا) على العبارات التالية:

*. إحدى الطرق المستخدمة لتحسين صدق الإختبار هي تطبيقه على مجموعة كبيرة من الأفراد.

*. يستخدم صدق المحتوى عند تطوير مقاييس الشخصية.

*. الصدق التدريسي هو شكل من أشكال صدق المحك التزامني.

*. الصدق الشكلي هو الذي يعكس مدى انسجام فقرة المقياس وفقا لتعريف باني أو مطور المقياس.

*. لا تتعدى قيمة الصدق الذاتي الجذر التربيعي لمعامل ثبات المتنبىء.

*. لا ينصح باستخدام المحكات الفورية في أغراض التنبؤ.

*. إن استخدام مجموعات صغيرة من المفحوصين في تقدير صدق المحك التنبؤي أو التزامني يؤدي إلى تضخيم الخطأ العيني.

*. إن إطالة الفترة الزمنية بين تطبيق المتنبىء وتطبيق المحك يساعد في فعالية عملية اتخاذ القرار.

*. إن إطالة الفترة الزمنية بين تطبيق المتنبىء وتطبيق المحك يساعد في تحسين صدق المحك التنبؤي.

س2: ضع دائرة حول رمز الإجابة الصحيحة للفقرات التالية:

*. أكثر أنواع الصدق أهمية بالنسبة لمقياس النزاهة هو ----.

أ) المحتوى

ب) التثبت

ج) البناء

د) التزامني

*. أفضل ما يمكن من الحكم على صدق اختبار في مجال أكاديمي من خلال ———

أ) مدى مطابقته لهدف المساق

ب) فعالية التعلم التي ينوي تحقيقها

ج) مدى مطابقته لما تم تعلمه في المساق

د) مستوى نوعية بناء فقراته

*. عندما نهتم بهدف أداة القياس، فإننا نعمل على التأكد من—————.

أ) صدق المحك التزامني

ب) صدق المحتوى

ج) صدق البناء

د) جميع ما ذكر

*. أكثر أنواع الصدق أهمية بالنسبة للإختبارات التحصيلية هو ————.

أ) المحتوى

ب) التنبؤي

ج) البناء

د) التزامني

*.إذا كان معامل ثبات أداة قياس يساوي (.33)، فإن ذلك يخبرنا بأن صدق الأداة ....

أ) جيدا للغاية

ب) ربما يكون جيدا

ج) غير جيد

*.إذا كان معامل ثبات أداة قياس يساوي (80.)، فإن ذلك يخبرنا بأن صدق الأداة ..

أ) جيدا للغاية

ب) ربما يكون جيدا

ج) غير جيد

س3: هل الإختبار الثابت صادقا ؟ كيف ؟

س4: أي نوع من أنواع الصدق الأكثر أهمية لكل من الإختبارات التالية:

اختبار ذكاء، اختبار فني (أدبي)، اختبار تحصيلي، قائمة استبيان للميـول، قائمة استبيان الاتجاهات؟

س5: إذا أردت أن تستخدم اختبار استعداد دراسي للجبر يساعدك في تحديد أسماء الذين سيأخذون فصلاً دراسياً في مادة الجبر للصف التاسع. وكان لديك اختبارين (A و B) متساويين في كل شيء (الكلفة، الشكل، وسهولة التصحيح) ماعدا ثباتهما وصدقهما التنبؤي كما في الجدول التالي، فأيهما تستخدم ؟

	الإختبار A	الإختبار B
الثبات	.84	.95
الصدق	.85	.80

الفصل السابع

المعايير والعلامات المعيارية

Norms and Normal Scores

- مقدمة
- أدوات القياس حسب طريقة تفسير مستوى الأداء
- تصنيف المعايير
- شروط تقييم الجماعة المرجعية
- معايير الأدوات معيارية المرجع
- نماذج من الدرجات المعيارية المحولة خطيا
- الدرجات المعيارية والرتب المئينية
- نماذج من بيانات معيارية
- الصفحة
- ملخص
- أسئلة وتمارين

7

المخرجات التعليمية

بعد دراسة هذا الفصل يتوقع من الطالب أن يكون قادرا على أن

- يقارن بين مفهومي المحك والمعيار.
- يقارن بين مفهوم الأدوات معيارية المرجع ومحكية المرجع.
- يصنف المعايير حسب الجماعة المرجعية.
- يقارن بين عمليتي التحويل الخطي وغير الخطي.
- يصف الدرجة المعيارية الزائية وكيفية حسابها.
- يصف الدرجة المعيارية التائية وكيفية حسابها.
- يصف العلاقة بين الدرجات المعيارية والمئينات.
- يصف الدرجات المعيارية التساعية وعلاقتها بالمئينات.
- يصف كيفية بناء معايير العمر الزمني.
- يصف كيفية بناء معايير الفرق الدراسية.
- يصف مفهوم الصفحة النفسية.

مقدمة

يستخدم المعيار **Norm** كمصطلح في القياس النفسي ليشير إلى متوسط درجات جماعة معينة من الأفراد على أحد المقاييس النفسية، ويطلق على تلك الجماعة (الجماعة المعيارية، أو المرجعية). والمعيار ضروري في القياس النفسي واختبارات التحصيل، لأن درجة الفرد التي يحصل عليها على الإختبار أو المقياس والتي تعرف بالدرجة الخام **Raw Score** ، ليس لها معنى بحد ذاتها، ولا تصلح للمقارنة مع درجته في مقاييس أو اختبارات أخرى، أو مع درجة شخص آخر على نفس المقياس إلا بواسطة المعايير (**Magnusson, 1967**).

والمعروف نظريا أنه لا معنى لأي مقدار كمي أو رقم دون معرفة المجموعة التي ينتمي إليها هذا المقدار. إذ لا يمكن معرفة أن الرقم (100) كبير أو صغير، دون تفحص موقعه بالنسبة للمجموعة التي ينتمي إليها ذلك الرقم. وبنفس الطريقة، فإنه لا يمكن تفسير درجة تلميذ على مقياس معين، إلا إذا تم مقارنتها بتوزيع درجات مجموعة مرجعية معينة ينتمي إليها التلميذ (**Glass and Hopkins, 1996; Glassnapp and Poggio, 1985**). فمثلا، إذا قيل بأن درجة طالب ما على امتحان في مادة القياس هي (75) فإنه لا يمكن تفسير مستوى درجة ذلك الطالب بشكل صحيح دون مقارنة درجته مع درجات زملائه بالصف وعلى نفس الامتحان. فربما تشير هذه الدرجة الخام (75) إلى أن هذا الطالب متميز أو ضعيف أو ربما راسب.

وفي مثال آخر، لو أن طالبا حصل على الدرجة (15) في اختبار المثابرة، فإن هذه الدرجة لا تعني شيئا بحد ذاتها، فهي غير قابلة للتفسير. ولكن لو عرفنا أن الدرجة الكلية في اختبار المثابرة هي (20)، عندها يمكن القول بأن الدرجة (15) تمثل (75 %) من الدرجة الكلية. وهذا التفسير لن يكون كاملا إلا إذا عرفنا أن متوسط درجات المجموعة التي ينتمي إليها

الطالب هي (14)، لأن ذلك يشير إلى أن هذا الطالب أفضل من متوسط مجموعته بدرجة واحدة.

ويجب أن نميز بين المحك **Criterion** أو المستوى القياسي **Standard** من جهة، والمعيار **Norm** من جهة أخرى. إذ أن المعيار يخبرنا عن الأداء الحقيقي للأفراد على الإختبار، بينما يشير المستوى القياسي إلى المستوى الذي يجب أن يصله أولئك الأفراد لكي تعتبر إجابتهم مقبولة. فمثلا تمثل الدرجة (100) المستوى القياسي للصحة النفسية في مقياس يقيس القلق ويتألف من (100) فقرة، وهذا غير ممكن من الناحية النظرية. غير أننا في الحياة الواقعية لا نجد شخصا ليس لديه شيء من القلق، لذلك نجد أن المعيار قد يكون في مثل هذا الإختبار (80). وبناءا على ذلك، فإن الشخص الذي يحصل على الدرجة (80) فأكثر يعتبر سويا لأنه يشبه أكثرية الجماعة التي ينتمي إليها (الزوبعي وبكر والكناني، 1981).

أدوات القياس حسب طريقة تفسير الدرجة

يعتمد تفسير الدرجة التي يحصل عليها الفرد على مقياس أو اختبار ما على نوع المرجعية التي يستند إليها في ذلك. فقد يتم تفسير درجة الشخص بالمقارنة مع متوسط أداء المجموعة التي ينتمي إليها ذلك الشخص. وتعرف هذه المجموعة بالجماعة المرجعية أو المعيارية (**Norm Group**)، وتسمى أدوات القياس التي يتم تفسير مستوى الأداء عليها بهذه الطريقة بالأدوات معيارية المرجع **Norm-Referenced -Tools (NRT)**. أما إذا تم تفسير درجة الشخص على أداة القياس اعتمادا على محك أداء متوقع (**Criterion**)، عندها تعرف تلك الأدوات بمحكية المرجع **Criterion-Referenced-Tools (CRT)**. ويتحدد هذا المحك بناءا على مجموعة المعارف والمهارات والكفايات التي يتوقع من الفرد امتلاكها. وفي هذا النوع من الأدوات يتم التعرف على ما يعرفه أو يتقنه وما لا يعرفه أو لا يتقنه الشخص من معلومات أو مهارات، بصرف النظر عن أداء المجموعة

التي ينتمي إليها. فعندما تتجه النية مثلا لتعيين طبيب جراح في مستشفى ما، فإن الحكم على المتقدمين يتم في ضوء مستوى إتقان كل منهم لمجموعة المهارات التي يتطلب المستشفى تحققها لدى الطبيب الذي سيتم تعيينه. عند ذلك، فإن المفروض أن يتم استبعاد كل طبيب لا يحقق النجاح في مجموعة المهارات المحددة، ولا بأس أن يتم ترتيب المتقدمين في ضوء معايير أخرى واختيار الأفضل للتعيين (Nunnally, 1978). وتعتمد محكات الأداء في الأدوات مرجعية المحك على النسب المئوية من المفردات التي تقيس سلوكا معرفا تعريفا دقيقا والتي يجيب عليها الفرد إجابة صحيحة.

تصنيف المعايير

تختلف معايير الأدوات معيارية المرجع باختلاف الجماعة المرجعية، الأمر الذي يتطلب ضرورة تحديد خصائص تلك الجماعة بأسلوب دقيق، كي يساعد ذلك في التفسير الدقيق للدرجات على الأداة. فقد يختلف ترتيب الشخص على الإختبار الواحد، إذا تم مقارنة درجته بدرجات مجموعات معيارية مختلفة. إذ ربما يكون ترتيب الشخص ضعيفا إذا ضمت الجماعة المرجعية متخصصين في موضوع الإختبار، وسيكون ترتيبه متقدما إذا كانت الجماعة المرجعية تتألف من طلبة مبتدئين أو متدربين في ذلك الموضوع.

وتنبثق أهمية تحديد الجماعة المرجعية من كونها أساسا في تفسير مستوى أداء الفرد، فهو ربما ينتمي لأكثر من جماعة مرجعية. فمثلا، إذا جلس شخص معاق سمعيا، عمره (15) سنة، وملتحق بالصف الثامن، لامتحان مقنن للمفاهيم الرياضية وحصل على الدرجة الخام (32) ، فقد تكون الرتبة المئينية لتلك الدرجة تساوي (99) إذا كانت الجماعة المرجعية للتفسير هي "مجموعة المعاقين سمعيا من طلبة الصف الثامن في الدولة"، أو تكون الرتبة المئينية لهذه الدرجة (96) إذا كانت الجماعة المرجعية هي

"طلبة الصف الثامن في منطقة تعليمية معينة" ، أو الرتبة المئينية (80) في جماعة "الملتحقين في مساق متقدم لتعلم الرياضيات"، أو الرتبة المئينية (78) في جماعة "أفراد مدرسة لرعاية المعاقين سمعيا". من هنا نفهم أن معـايير أداة القياس تختلف حسب الجماعة المرجعية التـي ينتمـي إليهـا الفـرد. والجدير بالذكر أن المعايير يمكن أن تصنف إلى وطنيـة، أو محليـة، أو حسب مؤسسة معينة، أو تبعا للفئات المستفيدة من الأداة (Nitko, 1983) كما هو موضح تاليا:

- **معايير وطنية National Norms** وتستخدم فـي حالـة تطبيـق أداة القياس على أفراد يتوزعون على مستوى الدولة أو الإقليم. وينطبق هذا في غالب الأحيان على اختبارات التحصيل والاستعدادات التي تستخدم على نطاق وطني. وخير مثال على ذلك اختبار الاسـتعداد المدرسي Scholastic Aptitude Test (SAT) الذي يـستخدم علـى نطاق واسع في الولايات المتحدة الأمريكية نهاية المرحلة الثانوية، بقصد التنبؤ بالنجاح بالدراسة الجامعية.

ويقوم بتحديد هذا النوع من المعايير هيئة وطنيـة مركزيـة مـن المحكمين والخبراء والسياسيين. ويراعى في عملية اختيـار عينـة تقنين المقياس أسس المعاينة لضمان التمثيل الأمثل للمجتمـع مـن ناحية، وتقليل الأخطاء إلى أدنى مستوى ممكن من ناحية أخرى.

معايير محلية Local Norms وتبرز الحاجة للاهتمام بالمعـايير المحلية عندما يكون هناك تباين واضح بين أفراد الجماعـة المرجعيـة الوطنية بسبب اختلاف البعد الجغرافي أو وجود عوامل أخرى. وذلـك عندما تكون أداة القياس موجهة إلى جماعة معينة تعيش فـي منطقـة معينة داخل الوطن. وخير مثال على ذلك، المعايير التي يتم وضـعها عندما يتم بناء اختبار لقياس مستوى تحصيل أبناء منطقة تعـاني مـن

ظروف اقتصادية معينة كما في المناطق النائية. والجدير بالذكر أن إعداد مثل هذه المعايير قد لا يحتاج إلى جهود متخصصة كما في حالة إعداد المعايير الوطنية. بل تعتمد على خلاصة مؤشرات تنموية لمنطقة معينة.

- **معايير تتعلق بالمؤسسة Institution Norms** كتلك التي تضعها مدرسة مهنية لقبول طلبتها على ضوء اختبار قبول ما، أو معايير تضعها مؤسسة صناعية أو تجارية للأدوات (الإختبارات والمقاييس) التي سيتم استخدامها لانتقاء المتقدمين لشغل وظيفة معينة.
- **معايير تتعلق بفئات معينة Special Norm Groups** كالمعايير التي يتم وضعها لاختبارات تتعلق بتصنيف المتفوقين دراسيا، أو تلك التي تستخدم في اختبارات الاستعدادات الموسيقية والجمالية، أو للاختبارات الموجهة لذوي الحاجات الخاصة، كالمعاقين عقليا أو حركيا عند توجيههم للقبول في برامج تأهيل خاصة. وعند زيارتك إلى أحد مراكز رعاية وتأهيل أصحاب الحاجات الخاصة، تجد أنهم يستخدمون اختبارات معينة لتشخيص وتصنيف حالات الإعاقة، وتقدير شدة كل منها للنزلاء اعتمادا على معايير معينة لكل اختبار.

شروط تقييم الجماعة المرجعية

جاء في تقرير جمعية علم النفس الأمريكية (APA, 1974) ضرورة توفر ثلاثة شروط أساسية في الجماعة المعيارية المرجعية **Norm Group** لتكون مناسبة للمعايير التي تستمد منها والتي يجب أن يصفها دليل أداة القياس **Instrument Manual** هي:

- **الحداثة Recency** ، فقد تتغير خصائص عناصر المجموعة المرجعية من وقت إلى آخر، نظرا لتأثير متغيرات بيئية أو ثقافية أو اجتماعية كما في القرارات والسياسات التربوية المتعلقة بالمناهج وطرق التدريس

وتقويمها. الأمر الذي يستوجب أن تكون المعايير مواكبة للتغيرات، وأن توثق تواريخ نشر أداة القياس في كل طبعة (Nitko, 1983; Allen and Yen, 1979). وتبرز أهمية هذا الشرط في ضوء ما يحدث من تغيرات متسارعة ومتشابكة تطرأ على خصائص المهن والوظائف وأساليب تقييمها. الأمر الذي يشير إلى ضرورة تقنين المقاييس عموما بين وقت وآخر في المجتمع الواحد بهدف تحديث جداول المعايير.

- **المواءمة Relevance** إذ يجب أن تكون خصائص الجماعة المرجعية موائمة لخصائص الفئة المستهدفة في تطبيق أداة القياس، خاصة عندما تكون الفئة المستهدفة هي مجموعة مهنية. وفي حال استخدام أداة القياس لأكثر من غرض، عندها ينبغي أن تستمد معاييرها من أكثر من جماعة مرجعية (Nitko, 1983).
- **التمثيل Representative**، إذ المفروض أن يتم اختيار أفراد الجماعة المرجعية بطريقة تضمن التمثيل الجيد للمجتمع المستهدف، بغض النظر عن عدد أفراد تلك الجماعة. ويفضل استخدام العينات العشوائية البسيطة إذا كان المجتمع متجانسا. وبغير ذلك، يفضل استخدام العينة العشوائية الطبقية عندما يكون الفرد وحدة للاختيار، أو العينة العشوائية العنقودية عندما تكون المجموعة كالصف الدراسي أو المستشفى هي وحدة الاختيار. ذلك بقصد ضمان أعلى درجات تمثيل خصائص المجتمع(Crocker and Algina, 1986; Nitko, 1983). وبشكل عام، تختلف طريقة المعاينة باختلاف التصميم الذي يتبع في تصميم الدراسة التي تستخدم فيها أداة القياس.

أنواع معايير الأدوات معيارية المرجع

عرفنا أنه ليس من الحكمة أن نقول بأن درجة طالب ما في مادة الرياضيات (85) أفضل من درجته في اللغة الإنجليزية (75)، ذلك لأن كل درجة مأخوذة من جماعة معيارية معينة، ومحسوبة حسب معيار معين. فقد

يختلف ترتيب الطالب في مادة اللغة الإنجليزية عنه في مادة الرياضيات، وقد تكون الدرجة النهائية في مادة الرياضيات مختلفة عنها في مادة اللغة الإنجليزية. هذا كله بسبب اختلاف مؤشرات النزعة المركزية والتشتت لتوزيع كل مجموعة من الدرجات. كما يشير أيضا إلى أنه من غير الصحيح أن تستخدم الدرجات الخام في وصف الأداء وكذا المقارنة.

ولمواجهة مثل هذه المشكلة (تفسير درجة ضمن مجموعة واحدة، أو مقارنة درجتين أو أكثر كل منها ينتمي إلى مجموعة مختلفة)، لا بد من إجراء عملية تحويل **Transformation** إلى شكل يمكن من خلاله إجراء مقارنة ذات معنى، وهناك نوعان من التحويل: خطي **Linear** ، وغير خطي **Non-Linear**. ويكون التحويل خطيا بإضافة ثابت إلى أو طرح ثابت من كل قيمة من قيم (درجات) التوزيع. أو ضرب كل قيمة أو قسمتها على ثابت، أو إجراء العمليات المذكورة مجتمعة. ولا يغير هذا النوع من التحويل من شكل توزيع البيانات الأصلي. ويدعى هذا التحويل بالخطي لأن العلاقة بين الدرجات الأصلية والدرجات المحولة تتخذ خطا مستقيما. بينما يكون التحويل غير خطي عند تحويل البيانات الأصلية إلى جذورها أو لوغاريتمها أو مقلوبها، ويعرف هذا النوع من التحويل بالتحويل غير الخطي لأن العلاقة بين البيانات الأصلية والبيانات المحولة ليست خطية. أي أن شكل توزيع البيانات بعد تحويلها تحويلا غير خطي يختلف عن شكلها قبل التحويل **(Besag and Besag, 1985; Blommers and Forsyth, 1977; Ferguson, 1966)**.

معايير الدرجات المحولة

إن أكثر أنواع الدرجات المحولة خطيا انتشارا هي: الزائية والتائية، والتساعية. والجدير بالذكر أنه يتم حساب كل من العلامة الزائية والتائية باستخدام معادلة رياضية معينة. وهناك علاقة رياضية تربط بين الدرجتين

الزائية والتائية وعلاقة كل منها بالتوزيع الطبيعي (**Glass and Hopkins, 1996**). وسيرد شرح مفصل عن كيفية حساب وتفسير كل من هذه الدرجات تاليا.

الدرجة الزائية

تعرف الدرجة المعيارية الزائية **z-score** بمقدار انحراف الدرجة الخام (x) عن المتوسط الحسابي لتوزيع مجموعة الدرجات التي تنتمي إليها الدرجة الخام مقدرا بوحدات الانحراف المعياري لذلك التوزيع (**s**)، أي أن:

$$Z = \frac{X - \overline{X}}{S}$$

والدرجة الزائية هي درجة معيارية في توزيع متوسطه الحسابي يساوي (صفر) وإنحرافه المعياري يساوي (1). بذلك يمكن مقارنة الدرجة مع غيرها من الدرجات التي تنتمي لنفس التوزيع، وذلك بحساب الدرجة المعيارية الزائية لكل منها ومن ثم يكون التفضيل في ضوء مقدار قيمة الدرجة المعيارية (z). وبنفس الطريقة يمكن مقارنة الدرجات الخام في أكثر من مادة بعد تحويل كل منها إلى ما يناظرها من درجة معيارية زائية. ويمكن تفسير قيم الدرجة الزائية كما في الحالات الرئيسة التالية:

الدرجة المعيارية الزائية (1) تقابل الدرجة الخام (الأصلية) التي تزيد عن متوسط درجات المجموعة التي تنتمي إليها بمقدار انحراف معياري واحد. ومن مراجعة جدول التوزيع الطبيعي، يمكن القول بأن هذه الدرجة يقع تحتها (أو هي أفضل من) 84.13 % من مجموعة الدرجات.

- الدرجة المعيارية الزائية (-1) تعني أن الدرجة تقل عن متوسط مجموعتها بمقدار انحراف معياري واحد. ومن جدول التوزيع الطبيعي، يمكن القول بأن هذه الدرجة أفضل من 15.87 % من مجموعة الدرجات.

- o الدرجة المعيارية الزائية (0) تعني أن الدرجة تساوي متوسط درجات المجموعة التي تنتمي إليها. ويقع تحتها نصف الدرجات أو هي أفضل من 50 % من مجموعة الدرجات.

وإذا أردنا تفسير الدرجة (65) التي حصل عليها طالب في امتحان مادة التاريخ. وكان المتوسط الحسابي لدرجات طلبة الصف في هذه المادة يساوي (75) والإنحراف المعياري لتلك الدرجات يساوي (5)، فإنه يمكن حساب الدرجة المعيارية الزائية لدرجة ذلك الطالب وتساوي (2-). وهذا بدوره يعني أن أداء الطالب في مادة التاريخ ضعيفا. إذ أن درجته تقل عن متوسط أداء مجموعته بمقدار انحرافين معياريين، أو أنها أفضل من (2.28 %) منهم فقط.

لمقارنة أداء طالب ظهرت درجتاه في اختبارين، أحدهما يقيس القدرة العددية والآخر يقيس القدرة اللغوية بواقع (85) و (75) على الترتيب، يمكن استعراض درجته الزائية (z) المقابلة لكل درجة في كل اختبار على النحو التالي:

	القدرة العددية	القدرة اللغوية
درجة الطالب	85	75
متوسط درجات الطلبة	70	70
الانحراف المعياري	10	5

وباستخدام معادلة الدرجة الزائية، تكون الدرجة الزائية لذلك الطالب في اختبار القدرة العددية (1.5) ودرجته الزائية في اختبار القدرة اللغوية (1). الأمر الذي يشير إلى أن أداءه على اختبار القدرة العددية أفضل من أداءه على اختبار القدرة اللغوية. وبشكل عام، تتصف الدرجة الزائية بالخصائص الرئيسة التالية:

- الفرق بين أي درجتين زائيتين في أي موقع في التوزيع يقابل فرقـا ثابتا في الدرجات الخام.
- مجموع الدرجات الزائية يساوي صفرا، والإنحراف المعياري لها يساوي (1).
- مجموع مربعات الدرجات الزائية يساوي عدد قيم عناصر العينة (n)، ويساوي عدد عناصر أفراد المجتمع ناقصا واحد (N-1).
- تحوي الدرجات الزائية قيما موجبة وسالبة وكسورا. الأمر الـذي لا يكون يسيرا على أولياء الأمور، أو العامة من النـاس اسـتخدامها أو تفسيرها.

الدرجة التائية

للتخلص ما أمكن من التعامل مع القيم السالبة والكسرية للـدرجات الزائية، يتم اللجوء إلى تحويلها إلى درجات معيارية أكثر عملية وسـهولة تعرف بالدرجة التائية **T-score**. وذلك بضرب الدرجة الزائية بمقدار (10) وإضافة (50) للناتج كما في المعادلة التالية:

$$T = 10Z + 50$$

والدرجة التائية هي درجة معيارية في توزيـع متوسـطه الحـسابي يساوي (50) وانحرافه المعياري يساوي (10). وتمتد القيم التائية للتوزيـع الواحد بين (صفر) و (100). ولما كان (99 %) من الدرجات تقـع ضـمن ثلاثة انحرافات معيارية في التوزيع الإعتدالي، فإنـه لا يتوقـع وجـود درجات متطرفة خارج خمسة انحرافات معيارية. وللحصول على الدرجـة التائية لدرجة خام، لا بد من تحويلها أولا إلى درجة معيارية زائيـة (z)، ومن ثم التعويض في المعادلة.

مثال: ما الدرجة التائية التي تقابل الدرجة الزائية (z =2.5) ؟

الحل:

$T = 10z + 50$;

$T= 10 (2.5) + 50$; $T =75$

مثال: ما الدرجة الزائية التي تقابل الدرجة التائية **90 (T = 90) ؟**

الحل:

$T = 10 z + 50$;

$90 = 10 z + 50$;

$10 z = 90-50$;

$z = 4$

مثال: ما الدرجة التائية التي تقابل الدرجة الزائية (2.2- = z)؟

الحل:

$T = 10 (z) + 50$;

$T = 10 (-2.2) + 50$;

$T = 28$

مثال: ما الدرجة الزائية المقابلة للدرجة التائية 10 (T = 10) ؟

الحل:

$Z = (T - 50) \div 10$;

$Z = (10 - 50) \div 10$;

Z= - 4

يبين الجدول التالي عرضا لتوزيع الدرجات الخام (x) لعينة من عشرة طلاب، وما يقابل كل منها من الدرجات الزائية (z) والتائية (T):

توزيع درجات عشرة طلاب وما يقابلها من القيم الزائية والتائية

الدرجة التائية (T)	الدرجة الزائية (z)	الدرجة الخام (x)
62.9	1.29	10
59.7	0.97	09
40.3	-0.97	03
62.9	1.29	10
59.7	0.97	09
37.1	-1.29	02
37.1	-1.29	02
62.9	1.29	10
46.8	-0.32	05
46.8	-0.32	05
20.0	20.0	n = 20
50.0	0.00	$\bar{x} = 60$
10.0	1.00	S = 3.1

يلاحظ من استعراض البيانات في هذا الجدول أن المتوسط الحسابي لقيم الدرجات الزائية والتائية هو (صفر) و (50) على الترتيب. وأن الانحراف المعياري لكل منهما هو (1) و (10) على الترتيب.

الدرجة التساعية

الدرجة التساعية Stanine هي درجة معيارية في توزيع يتألف من تسع فئات، كل منها يسمى تساعي ويضم نسبة معينة من القيم (درجات أو صفات). ولا يخضع هذا التوزيع إلى قاعدة رياضية أو قانون إحصائي، بل

يستند إلى معطيات تعتمد على مصلحة المؤسسة المستفيدة من هذه الدرجات (رمضان، 1981؛ عنبر، 1995).

ويتم إيجاد الدرجة التساعية عن طريق تحويل كل درجة خام إلى رتبة مئينية، ومن ثم تصنيف الرتب المئينية في فئات، حيث أن كل واحدة تقابل تساعيا معينا ويقابل كل تساعي نسبة معينة من القيم كما في الجدول التالي:

القيم التساعية وما يقابلها من نسب مئوية

التساعي	1	2	3	4	5	6	7	8	9
النسبة	4	7	12	17	20	17	12	7	4
النسبة التراكمية	4	11	23	40	60	77	89	96	100

ويمكن تحويل الدرجات الخام إلى درجة زائية، ومن ثم يتم إيجاد رتبتها المئينية التي تصنف في فئات، والتي يقابل كل منها تساعيا وهكذا. وبالنظر إلى الجدول، يتضح أن أعلى 4% من الدرجات تقع في التساعي التاسع وأدنى 4 % تقع في التساعي الأول. كما يتضمن كل من التساعي الثاني والثامن 7 % من الدرجات ، وأن 77 % من الدرجات تقع دون التساعي السادس.

مثال: إذا كانت درجة "بسام" في امتحان للإحصاء (75) في توزيع متوسطه الحسابي (65) وانحرافه المعياري يساوي (5)، فأوجد درجته: الزائية، والتائية، والتساعية، إذا افترض أن توزيع الدرجات التي تنتمي إليه درجة "بسام" طبيعيا.

الحل:

$$z = (75 - 65) \div 5 = 2$$

$$T = 10 \times 2 + 50 = 70$$

من جدول التوزيع الطبيعي ، نجد أن الاحتمال الذي يقابل الدرجة الزائية يساوي **98 %** وتقع هذه النسبة ضمن التساعي التاسع.

معايير المئينات والرتب المئينية

تبرز أهمية الحديث عن المئينات عندما نريد أن نبين موضع شخص ما أو قيمة معينة بالنسبة لجماعة معروفة. والنقطة المئينية **percentile point** هي تلك الدرجة التي يقع تحتها نسبة مئوية محددة من الحالات في المجموعة. وهذه النسبة المئوية ما تعرف بالرتبة المئينية **Percentile rank** **(Ferguson, 1966)**.

المئين والرتبة المئينية مصطلحان مختلفان من الواجب التفريق بينهما، فالرتبة المئينية نسبة مئوية تأخذ قيما تتراوح بين (صفر) و **(100)**، بينما يعتبر المئين نقطة أو علامة يمكن أن تأخذ أي قيمة، فمثلا قد تكون علامة طالب في إمتحان **GRE** هي **(870)** وتكون رتبته المئينية في ذلك الامتحان تساوي 64 % هذه الدرجة (870) يقع تحتها 64 % من درجات الطلاب الذين أخذوا ذلك الامتحان في تلك الفترة (**Gay, Mills, and Airasian, 2008; Weinberg, Sharon, Goldberg, 1990**).

كما ويمكن القول بأن قيمة المئين **64** لذلك الطالب تساوي **(870)**. وبلغة أخرى، يمكن التوضيح بين المصطلحين على النحو التالي: عندما يكون المطلوب حساب الرتبة المئينية للقيمة **(70)** يكون المطلوب إيجاد النسبة المئوية من مجموع القيم التي تقع تحت القيمة **(70)**. وعندما يكون المطلوب حساب المئين 70، يكون المطلوب إيجاد القيمة أو النقطة أو الحد الفعلي الأعلى الذي يقع تحته 70 % من القيم (النبهان، **2008**).

مثال: باستخدام البيانات الواردة في الجدول التالي، أحسب قيم كل من: P_{25} والرتبة المئينية للقيمة **(60)**:

الفئة	f
30 – 34	2
35 – 39	2
40 – 44	4
45 – 49	9
50 – 54	6
55 – 59	5
60 – 64	3
65 – 69	2
70 – 74	0
75– 79	1

الفئة	f	الحدود الفعلية العليا	Cf
30 – 34	2	34.5	2
35 – 39	2	39.5	4
40 - 44	4	44.5	8
45 – 49	9	49.5	17
50 – 54	6	54.5	23
55 – 59	5	59.5	28
60 – 64	3	64.5	31
65 – 69	2	69.5	33
70 – 74	0	74.5	33
75 – 79	1	79.5	34

الحل: ننظر إلى الفئات وإلى تكراراتها التراكمية Cf فقط، لــذلك من المفروض أن نكون عمودا للحدود الفعلية العليا للفئات وآخر للتكرارات التراكمية كما هو مبين تاليا:

المئين 25 (P_{25}):

$(25 \div 100) \times (34) = 8.5$

ننظر إلى عمود التكرار التراكمي ونعين موقع (8.5) على هذا العمود ويتبين أنها تقع بين التكرارين التراكميين (8) و (17)، ومن ثم نحسب ما يناظرها في عمود الحدود الفعلية العليا للفئات بطريقة النسبة والتناسب. أي أن المئين 25 يقع بين القيمتين (44.5) و (49.5)، والمطلوب الآن تحديد موقع هذا المئين بالضبط. ويتم ذلك بإجراء الحساب اللازم (النسبة والتناسب) لنجد أن المئين 25 = (44.77).

الرتبة المئينية للدرجة 60 (PR_{60}):

هنا ننظر إلى عمود الحدود الفعلية العليا من أجل البحث عن القيمة (60) والتي يظهر أنها تقع بين (59.5) و (64.5). والمطلوب هو إيجاد التكرار التراكمي المقابل لهذه الدرجة. يقع هذا التكرار بين القيمتين (28) و (31). وباستخدام النسبة والتناسب يمكن حساب التكرار التراكمي المطلوب ويساوي (28.3) وتكون الرتبة المئينية المطلوبة كما يلي:

$$(28.3 \div 34) * 100\ \% = 83\ \%$$

مما تقدم يمكن القول بأن الرتبة المئينية للدرجة (60) هي (83 %). أي أن هناك 83 % من الدرجات تقع دون الدرجة (60).

وجدير بالذكر أن هناك جداول تستخدم في إجراء مقارنات داخل الفرد وبين الأفراد، والجدول التالي يعرض بيانات عن درجات أحد الممتحنين على اختبارات بطارية الاستعدادات الفارقية (DAT) التي تقس عددا من القدرات الفرعية، ويقابل كل درجة رتبتها المئينية:

الدرجات الخام لأحد الممتحنين على اختبارات القدرات الفرعية في بطارية الاستعدادات الفارقية والرتبة المئينية المقابلة لكل منها

القدرة الفرعية	الدرجة الخام (x)	الرتبة المئينية PR
الاستدلال اللفظي	33	65
القدرة العددية	27	70
الاستدلال المجرد	34	45
الاستعداد الكتابي	41	40
الاستعداد الميكانيكي	40	20
العلاقات المكانية	35	55
القدرة الهجائية	62	55
القدرة النحوية	29	65

ويلاحظ من الجدول أن رتبة الطالب المئينية المناظرة لدرجته الخام في اختبار القدرة العددية (27) تساوي (70)، أي أن أداءه في ذلك الإختبار يتفوق على (70 %) من زملائه في هذه القدرة. في حين لم يتجاوز مستوى أدائه على اختبار الاستعداد الميكانيكي سوى (20 %) من أقرانه.

ويختلف تقييم مقدار الرتبة المئينية حسب الغرض من الإختبار ومدى قناعة الجهة التي تصمم الإختبار لتحقيق أهداف معينة. فقد تكون الرتبة المئينية لدرجة طالب على اختبار معين 80 غير مقبولة لأغراض القبول في برنامج دراسات عليا. في حين قد يبدو ذلك معقولا، إذا هدف الإختبار إلى تصنيف المتقدمين لشغل وظيفة خدمية مثلا. وبشكل مختصر، لا يمكن للرتبة المئينية أن تنبئنا عن مقدار ما يمتلكه الفرد من مهارات ومعلومات وخبرات تتعلق بمحتوى المقياس أو الإختبار. بل أنها تحدد مكانة الفرد بالنسبة لأفراد جماعته المرجعية. كما أنه من غير المعقول اعتبار الرتب المئينية قيما مطلقة أو غاية في الدقة، نظرا لأن الأفراد الذين يحصلون

على رتبتين مئينيتين متقاربتين مثل (51)، (52) تكون درجاتهم متقاربة. الأمر الذي حدا بناشري الإختبارات أن يدونوا المئينات في أدلة الإختبارات بحيث تتضمن مدى معين بدلا من التعامل معها كنقطة. فبدلا من القول بأن الدرجة الخام (65) في اختبار ما تناظر المئين (58) يمكن اعتبارها مناظرة للمدى المئيني (57– 60).

معايير الارتقاء (النمو)

تميز سلوك الإنسان بصفة النمو في مختلف السمات بمرور الزمن، وبطريقة منظمة. فقد يرتقي مستوى تحصيل الفرد وكذلك قدراته العقلية ومهاراته بانتقاله من مستوى صفي إلى آخر، ومن مستوى عمري إلى آخر. وعليه، فقد بات من الضروري أن يتم بناء معايير تستخدم في مقارنة أداء الفرد بمؤشرات أداء المجموعة التي ينتمي إليها، وكذلك مقارنة أدائه بمستويات نمائية متعددة. ومن أكثر معايير الارتقاء استخداما هي:

- معايير العمر الزمني
- معايير الصف الدراسي أو الفرق الدراسية

معايير العمر الزمني

يمكن تهيئة مجموعة من معايير العمر الزمني **Age Norms** لأية خاصية سلوكية تنمو بنمو الفرد، خاصة إذا افترضنا أن نمو الفرد يحدث نتيجة لعوامل النضج. وتسمى هذه المعايير أحيانا بالعمر التعليمي **Educational Age** أو مكافىء العمر الزمني **Age Equivalent**. فإذا كان سعيد طالبا في الصف الخامس الابتدائي وطوله (140) سم فربما نعتبره الأطول من بين زملائه في الصف الخامس. ولكن من الأفضل مقارنة طول سعيد مع أقرانه من نفس مستواه العمري، فإذا كان عمره (10) سنوات و (8) شهور، ووسيط الطول لأطفال جماعته المرجعية (الطلبة ذوو الوسيط العمري

(10) سنوات و (8) شهور) هو 142 سم، فإنه يعد أقصر من وسيط طول زملائه. بذلك فإنه من الممكن تقييم طول الفرد إذا توفرت لنا معلومات عن معايير العمر الزمني أو معايير مكافىء العمر للأطوال، لأن هذا يمكننا من تحديد الجماعة المرجعية المناظرة لخاصية الفرد موضع البحث.

ولتحديد مكافىء العمر لطول سعيد، ينبغي أن نحدد جماعة مرجعية وسيطها (140) سم. فإذا كان عمر أفراد هذه الجماعة (10) سنوات و (6) شهور، فإن مكافىء عمر سعيد يكون (10) سنوات و (6) شهور، ويكتب (10.6). وللحصول على مكافىء العمر في خاصية معينة من خصائص الفرد، يعطى الإختبار المتعلق بتلك الخاصية على مجموعات من الأفراد ممثلة لمستويات عمرية مختلفة، ثم نحسب قيمة وسيط توزيع الدرجات على الإختبار في كل مستوى عمري على حدة. ويمكن عندئذ تفسير الدرجات الخام عل أنها تمثل مستوى أداء مناظر لوسيط أداء عمر زمني معين.

ومهما يكن من أمر، فإن معيار العمر الزمني مناسب للخصائص والسمات السلوكية التي تظهر استمرارية ونموا مضطردا بتقدم العمر. وبالرغم من افتراض هذا المعيار لمقولة أن تلك الخصائص والسمات تنمو باتساق في مختلف الأعمار، إلا أنه قد يحدث أن يتغير معدل النمو أو ربما يتوقف في مستويات عمرية معينة.

ومن أبرز الصعاب التي تواجه المشتغلين في بناء معايير العمر الزمني أثناء بناء وتطوير أدوات القياس، الحصول على عينة كبيرة نسبيا وممثلة لكافة الفئات العمرية بشكل متوازن للتمكن من إيجاد مؤشرات تتسم بالثبات والاستقرار لكل مستوى عمري.

معايير الفرق الدراسية

يشيع استخدام هذا النوع من المعايير في تفسير درجات الأشخاص في الإختبارات عموما والتحصيلية على وجه الخصوص. ويطلق على هذا النوع من المعايير "معايير مكافىء الفرق الدراسية" **Grade Equivalent**.

وتناسب هذه المعايير الإختبارات التحصيلية التي يستخدمها التربويون في المرحلة الابتدائية غالبا. فالدرجة التي يحصل عليها طالب في الإختبار يتم مقارنتها بوسيط درجات الجماعة المرجعية في فرقة دراسية معينة.

ولحساب هذا النوع من المعايير، يتم تطبيق الإختبار على جماعة مرجعية ممثلة لمختلف الفرق الدراسية وإيجاد وسيط درجات كل فرقة على حدة. ويستخدم وسيط كل من هذه الفرق في تفسير الدرجات الخام، أي أننا نحدد الفرقة الدراسية المكافئة لأي درجة خام في الإختبار.

ويعبر عن مكافىء الفرقة الدراسية برقمين يفصل بينهما فاصلة، يشير الرقم الثاني إلى الصف ويشير الرقم الأول إلى ذلك الشهر من العام الدراسي. فالمكافىء (3.8) يشير إلى الفرقة الثالثة أي الصف الثالث والشهر الثامن من العام الدراسي.

وتعتمد هذه المعايير اعتمادا أساسيا على أساليب انتقاء المدارس والطلبة والسياسات التعليمية التي سبق أن أشرنا إليها، ويتم تدوين هذه المعايير على أساس عشرة أشهر وليس 12 شهرا، كما في معايير مكافىء العمر، أي تتراوح مثلا بين (4.0) و (4 .9) للفترات الشهرية للعام الدراسي الرابع. ويبين الجدول التالي الأجزاء الكسرية لمكافىء الفرق الدراسية:

الأجزاء الكسرية لمكافىء كل فرقة دراسية

الشهر	الجزء الكسري	الشهر	الجزء الكسري
أيلول/ سبتمبر	.0	شباط/فبراير	.5
تشرين أول/أكتوبر	.1	آذار/مارس	.6
تشرين ثاني/نوفمبر	.2	نيسان/ابريل	.7
كانون أول/ديسمبر	.3	أيار/مايو	.8
كانون ثاني/يناير	.4	حزيران/يونيو	.9

ويبين الجدول التالي نموذجا لمعايير مكافىء العمر الزمني ومعايير الفرق الدراسية لاختبار افتراضي في فهم معاني الكلمات كما اقتبس من علام (2000).

نموذج لمعايير كل من مكافىء العمر الزمني والفرق الدراسية لاختبار افتراضي

مكافىء الفرقة الدراسية	مكافىء العمر الزمني	الدرجة الخام
1.9	7.1	1
2.0	7.2	2
2.1	7.3	3
2.2	7.4	4
2.3	7.5	5
2.4	7.6	6
2.5	7.7	7
2.6	7.8	8
2.7	7.8	9
2.8	7.9	10
2.9	8.0	11

ويتضح من هذا الجدول أن الدرجة الخام (1) تناظر مكافىء العمر الزمني (7.1) سبعة أعوام وشهر واحد، كما تناظر مكافىء الفرقة الدراسية الأولى (1.9) أي الصف الأول والشهر التاسع من العام الدراسي. بينما تناظر الدرجة (10) مكافىء العمر الزمني (7.9)، أي سبعة أعوام وتسع شهور، كما تناظر مكافىء الفرقة الدراسية (2.8)، أي الصف الثاني والشهر الثامن، وهكذا.

وفي الحالات التي نريد أن نفسر الدرجات التي تقع خارج نطاق المنحنى أو الجدول المعطى (الدرجات العالية أو المنخفضة)، سواء في حالة المعايير العمرية أو الدراسية، فإنه ينبغي إجراء عملية استكمال **extrapolation** للجدول أو للمنحنى. ونظرا لسهولة بناء وفهم معايير الفرق الدراسية، فقد شاع استخدامها بين العاملين في التربية بشكل واسع في البلاد الأجنبية، وخاصة في المرحلة الابتدائية.

وبالرغم من انتشار استخدام معايير الفرق الدراسية، إلا أنه ليس دقيقا في تقييم التلاميذ وخاصة في الصفوف المتقدمة من الدراسة الابتدائية والإعدادية. وكثيرا ما يساء فهم هذا المعيار عندما يستخدمه المعلمون في الحكم على أن كل طالب يصل إلى الدرجة المتوسطة أو يزيد بأنه" طالب ناجح". وهذا قول غير دقيق، لأن هذه المعايير متوسطات. ومن الطبيعي أن يكون هناك طلبة لا تسمح لهم قدراتهم الوصول إلى المتوسط وبذلك فإنه لا يجوز من الناحية التربوية أن يطلق عليهم "متأخرون دراسيا" (الزوبعي، بكر، الكناني، **1981**).

كما قد يخطىء البعض في تفسير مكافىء الفرق الدراسية، فقد يفهم البعض أن المعيار يشير إلى الفرقة التي ينبغي للطالب أن يلتحق بها. فإذا كان مكافىء الفرقة الدراسية لطالب في الصف الرابع الابتدائي (7.4) فهذا لا يعني أنه يجب نقل هذا الطالب إلى الفرقة السابعة، أو أنه يجب إعطاؤه واجبات مدرسية تناسب الفرقة السابعة، ولكنها تعني فقط أن أداء هذا الطالب يفوق متوسط أو وسيط أداء طلبة الصف الرابع.

نماذج من الدرجات المعيارية المحولة خطيا

هناك عدد من الدرجات المعيارية المحولة خطيا، والتي تستخدم في أكثر من مجال في تفسير السلوك الإنساني كنسبة الذكاء الانحرافية **DIQ**، التي تمثل درجات معيارية ذات متوسط مقداره (100) وانحراف معياري مقداره **(16)**:

$$DIQ = 16z + 100$$

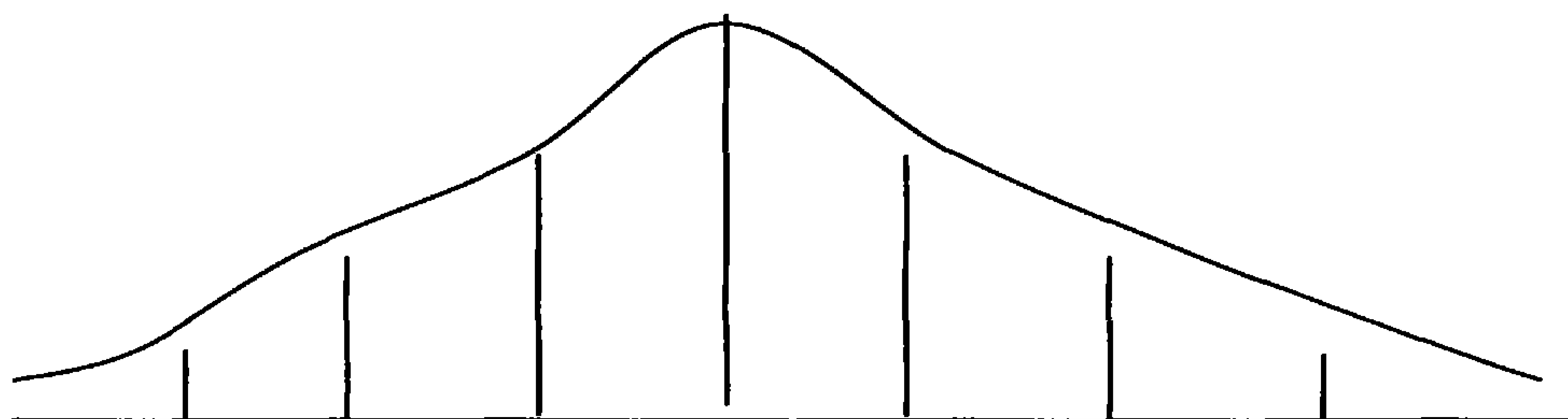
وكذلك درجات الامتحان **GRE** ذات المتوسط الحسابي (**500**) والانحراف المعياري = (**100**) التي تمثل درجات محولة كالدرجات التائية:

$$GRE = 100z + 500$$

إضافة إلى درجات اختبار الاستعداد الدراسي (**SAT**) التي تتوزع بمتوسط حسابي مقداره (**500**) وانحراف معياري (**100**):

SAT =100 z + 500

ويعرض الشكل التالي (1) نماذج لبعض الـــدرجات المعياريـــة المحولـــة والرتب المئينية المقابلة لها على افتراض أن التوزيع الأصـــلي للـــدرجات الخام يمثل توزيعا إعتداليا (Gay, Mills, and Airaisan, 2000):



Z	-3	-2	-1	0	+1	+2	+3
T	20	30	40	50	60	70	80
IQ	52	68	84	100	116	132	148
PR	.1	2	16	50	84	98	99.9
GRE	200	300	400	500	600	700	800
SAT	200	300	400	500	600	700	800

الشكل (1) نماذج لدرجات معيارية محولة خطيا

نلاحظ من هذا الشكل بأن الدرجة (500) على كـــل مـــن اختبـــار GRE واختبار SAT تقابل الدرجة التائية (50) والدرجة الزائية (صفرا) والرتبة المئينية (50).

الدرجات المعيارية والرتبة المئينية

مما تجدر الإشارة إليه أن الرتبة المئينية (PR) تعد درجـــة معياريـــة محولة ولكنها تقع على مقياس رتبي، بخلاف الدرجتين الزائيـــة والتائيـــة اللتين يقاس كل منهما على المستوى النسبي والفئوي. وهذا يحد من إجراء

بعض العمليات الرياضية مثل إيجاد الوسط الحسابي والانحراف المعياري على الرتب المئينية، بخلاف الدرجتين الزائية والتائية. إضافة إلى أن الفرق بين رتبتين مئينيتين في وسط التوزيع، يقابل فرقا في الدرجات الخام أقل من الفرق في الدرجات الخام المقابلة لنفس الفرق بين رتبتين مئينيتين في أي من طرفي (ذيلي) التوزيع. بينما يكون الفرق بين علامتين زائيتين في أي موقع في التوزيع فرقا ثابتا في الدرجات الخام.

مثال: إذا كانت الرتبة المئينية لدرجة "سعاد" في امتحان تحصيلي تساوي 77 فما الدرجة المعيارية الزائية والدرجة التائية المقابلة لدرجة "سعاد" ؟

الحل: الرتبة المئينية لدرجة "سعاد" تساوي (77)، بمعنى أن (77 %) من الدرجات يقع تحت تلك الدرجة. ولمعرفة مقدار الدرجة المعيارية الزائية المقابلة لتلك الدرجة، يستلزم الرجوع إلى جدول التوزيع الطبيعي لإيجاد قيمة الدرجة (z) التي تقع تحتها المساحة (.77) من المنحنى الطبيعي وكانت تساوي (.75) وبذلك تكون الدرجة التائية المقابلة لدرجة "سعاد" تساوي:

$$T = 10\,z + 50\ ;$$

$$T = 10\,(.75) + 50\ ;$$

$$T = 57.5$$

نموذج بيانات معيارية

تم اختيار نموذج من عينة من معايير الأداء على مقياس بيتا III للذكاء غير اللفظي المكون من خمسة اختبارات فرعية. وقد تم تطوير معايير هذا المقياس على عينة من طلبة جامعة الامارات العربية المتحدة في صيف عام 2002/2001. وسيتم عرض النموذج تاليا:

معايير نسب الذكاء على مقياس بينا III

تتيح معايير مقياس بينا III التعبير عن مستوى أداء الفرد الممتحن بدلالـــة نسبة الذكاء IQ والمئين. ويختلف هذا المقياس عن غيره من المقاييس في أن الدرجات المعايرة scaled scores تصحح بدلالة الفئة العمرية. وقد تـــم بناء هذا النوع من الدرجات بنفس الطريقة التي تمت في حالـــة مقـــاييس ويكسلر لذكاء الراشدين Wais-III (Kellogg & Morton, 1999).

تم اشتقاق المعايير بتحويل الدرجات الخام على كل اختبـــار فرعـــي إلـــى درجات معايرة بمتوسط مقداره (10) وانحراف معياري قدره (3). وتبـــع ذلك إيجاد مجموع الدرجات المعايرة الخمس، ثم جرى تحويله إلى توزيـــع درجات نسب الذكاء IQ بمتوسط مقداره (100) وانحراف معيـــاري قـــدره (15). وقد تم بناء جداول تحويل الدرجات الخام على كل اختبار فرعي إلى درجات معايرة كما في الجدول التالي:

وجدير بالقول هنا أن عملية جمع الدرجات المعايرة بدلا من الدرجات الخام تتيح القول بأن الإختبارات الخمس الفرعية متساوية في الوزن في تحديـــد نسب الذكاء IQ أو المئينات. كذلك فإن مـــن الحكمـــة أن لا يـــتم تفـــسير الدرجات على كل اختبار فرعي بشكل منفصل، بل يجب أن ينظر إلى أن درجة الذكاء الكلية IQ تحقق تفسيرات ذات معنى وتتمتع بثبـــات وصـــدق عاليين.

الدرجات المعايرة المكافئة للدرجات الخام (20– 24) سنة

استدلال المصفوفات	فرز الصور	فحص التطابق	إكمال الصور	الترميز	الدرجة المعايرة (z)
0-2	0-3	(−2) أو أقل	0-1	0-20	1
3-4	4-6	(-1) – 7	2-3	21-32	2
5-7	7-9	8-18	4-5	33-41	3
8-9	10-12	19-24	6-7	42-47	4
10-11	13-15	25-29	8-9	48-52	5
12-13	16-17	30-32	10-11	53-56	6
14-15	18-19	33-34	12-13	57-63	7
16	20	35-37	14-15	64-68	8
17	21	38-39	16	69-72	9
18-19	–	40-41	17	73-76	10
20	22	42-43	18	77-81	11
21	23	44	19	82-86	12
22	–	45-46	–	87-92	13
23	24	47	20	93-98	14
24	–	48	21	99-103	15
25	–	49-50	22	104-110	16
--	–	51	23	111-119	17
–	–	52-53	24	120-132	18
--	–	54-55	--	133-140	19

أما عملية تحويل مجموع الدرجات المعايرة على الإختبارات الخمس إلى درجة الذكاء IQ فتتم من خلال الجدول التالي:

قيم درجات الذكاء المكافئة لمجموع الدرجات المعايرة

مجموع الدرجات المعايرة	بيتا III (IQ)	الرتبة المئينية	مجموع الدرجات المعايرة	بيتا III (IQ)	الرتبة المئينية
5	48	0.0	51	100	50
6	49	0.0	52	102	55
7	50	0.0	53	103	58
8	52	.1	54	105	63
9	53	.1	55	106	66
10	54	.1	56	107	68
11	55	.1	57	108	70
12	56	.2	58	109	73
13	57	.2	59	111	77
14	58	.3	60	112	79
15	60	.4	61	114	82
16	61	.5	62	115	84
17	62	1	63	117	87
18	63	1	64	118	88
19	64	1	65	120	91
20	65	1	66	122	93
21	66	1	67	124	95
22	67	1	68	126	96
23	68	2	69	128	97
24	69	2	70	129	97
25	70	2	71	131	98
26	71	3	72	133	72
27	72	3	73	135	99
28	73	4	74	137	99
29	74	4	75	139	99.5
30	75	5	76	141	99.7
31	76	5	77	142	99.7
32	77	6	78	144	99.8
33	78	7	79	146	99.9
34	89	8	80	148	99.9
35	81	10	81	149	99.9
36	82	12	82	151	100
37	83	13	83	153	100
38	84	14	84	155	100
39	85	16	85	155	100
40	86	18	86	155	100
41	87	19	87	155	100

42	88	21	88	155	100
43	989	23	89	155	100
44	91	27	90	155	100
45	92	30	91	155	100
46	94	34	92	155	100
47	95	37	93	155	100
48	96	39	94	155	100
49	98	45	95	155	100
50	99	47			

مثال: لاستخراج درجة ذكاء طالب ما باستخدام مقياس بيتا للذكاء غير اللفظي، يتم تطبيق المقياس باختباراته الخمس على ذلك الطالب مع مراعاة الزمن المخصص لكل منها. ثم يجري حساب درجته الخام على كل اختبار فرعي. فإذا كان هذا الطالب ينتمي إلى الفئة العمرية (20- 24) سنة، يتم الرجوع إلى الجدول وتحديد درجته على كل اختبار فرعي وما يقابلها من درجة معايرة كما يلي:

الإختبار الفرعي	الدرجة الخام	الدرجة المعايرة
الترميز	80	11
إكمال الصور	20	14
فحص التطابق	48	15
فرز الصور	20	8
استدلال المصفوفات	23	14

من جدول الدرجات المعايرة، نجد مجموع تلك الدرجات: **11+14+15+8+14 = 62** ، ومن جدول قيم الذكاء الإنحرافية، نجد أن درجة الذكاء للطالب تساوي **(115)** ورتبته المئينية تساوي **(84 %)**.

الصفحات Profiles

تساعد المعايير في إيجاد أسلوب نعبر من خلالــه عــن درجــات مجموعة من الأفراد في اختبارات مختلفة. ويكون هذا التعبيــر بوحــدات عامة، تجعل المقارنة مباشرة. وبالرغم من أن المعــايير الــسابقة تــسهل مقارنة درجات الفرد في اختبارات مختلفة تقيس نفس الشيء، إلا أن أحــدا من هذه المعايير لا يمكننا من مقارنة درجات الفرد في اختبــارات تقــيس وظائف مختلفة.

وإذا تم تحويل درجات أي اختبار يقيس أي وظيفة إلى نــوع مــن المعايير ينسب الفرد إلى مستوى مجموعة دراسية معينة في الرياضــيات مثلا، وينسبه أيضا إلى مجموعة أعلى في اللغة الإنجليزية، وينسبه اختبار آخر إلى مجموعة أقل من الأولى في العلوم، تسهل المقارنة. بذلك يكــون للفرد الواحد أكثر من مجموعة دراسية حسب مستواه في المواد المختلفــة المذكورة أعلاه.

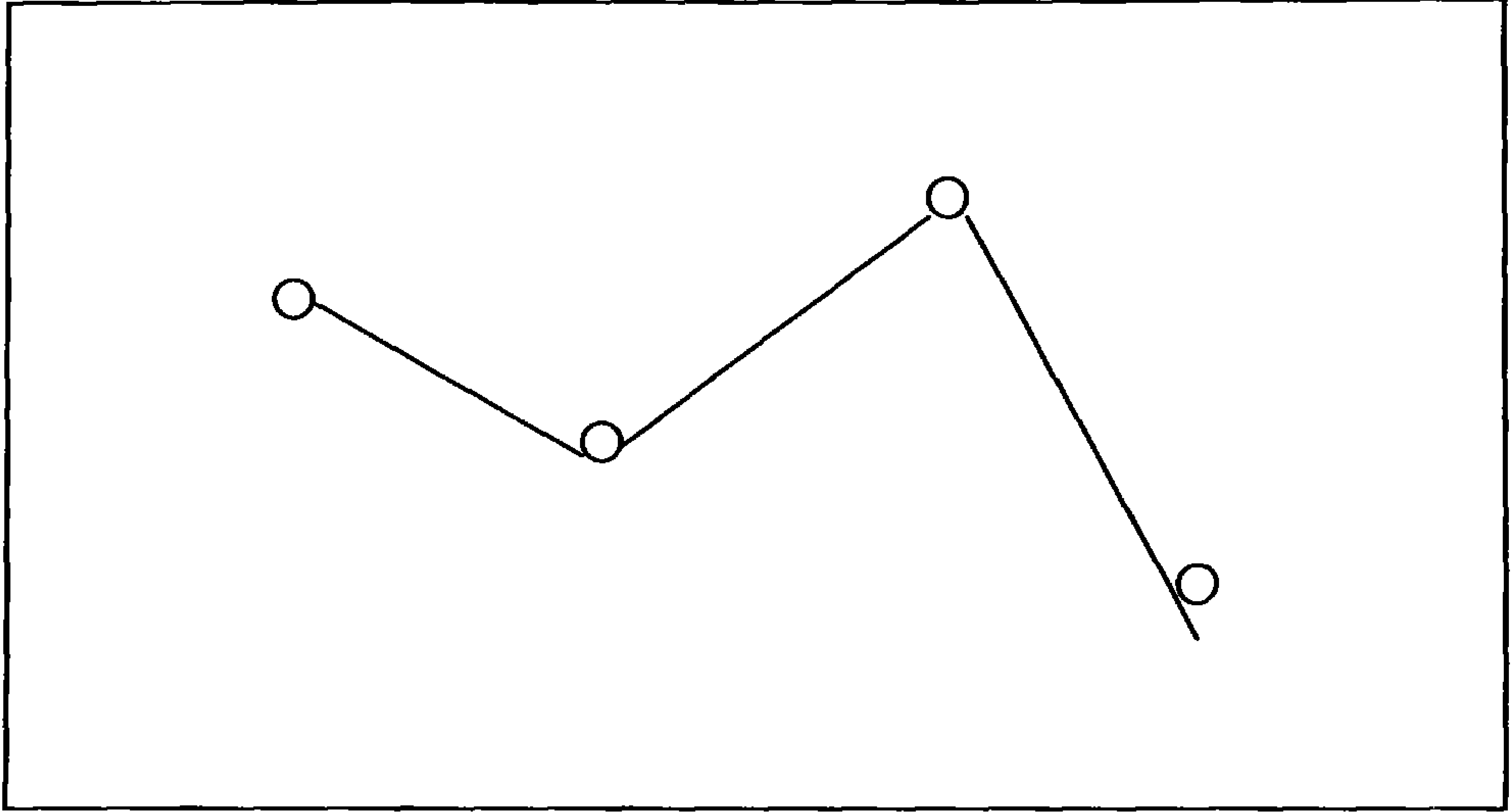
وقد أمكن جمع هذه المعايير في صفحة واحــدة تعــرف بالــصفحة النفسية **Psycho Profile**. وتأخذ الصفحات أشكالا وتصاميم مختلفــة. وقــد أمكن تعريفها بالتمثيل البياني النفسي لمستوى أداء الفرد على عــدد مــن الإختبارات أو مستوى أدائه على فقرات مقياس ما على مدى فترات زمنية متعاقبة. وتعرف الصفحة النفسية أحيانا بالخريطة النفسية أو التربوية.

والجدير بالذكر أن مفهوم الصفحة النفسية يرتبط بمفهــوم بطاريــة الإختبارات **Test Batteries**. إذ أن البطارية بمفهومها الأســاس، مقيــاس يتألف من اختبارات فرعية تقيس سمات أو قدرات أو مهارات. وتهدف إلى التعرف على مستوى أداء الفرد في كل اختبار فرعي من ناحيــة، وعلــى مقارنة مستوى أدائه في الإختبارات المختلفة. ومن أمثلة هذه البطاريات ، بطارية الاستعدادات الفارقة (**DAT**)، وبطارية الاستعداد الأكاديمي (**SAT**)،

وبطارية اختبارات الذكاء والشخصية مثل اختبار ويكسلر لذكاء الأطفال (WISC) أو اختبار ويكسلر لذكاء الراشدين (WAIS) وبطارية اختبار مينيسوتا الشخصية متعدد الأوجه (MMPI).

وتبنى الصفحة النفسية أو التربوية لمجموعة من درجات الإختبارات لفرد ما بتمثيل الإختبارات على المحور الأفقي، ويمثل مقدار كل درجة بنقطة على المحور العمودي. ثم يصار إلى توصيل النقاط بخط مستقيم أو بأعمدة بيانية، كما يتم تحديد قيم المئينات أو الدرجات المعيارية على المحور العمودي لتوضيح عملية المقارنة.

ويمكن تمثيل الصفحة النفسية بشكل بسيط، حيث يمكن تمثيل القدرات الفرعية على المحور الأفقي ثم تمثل الدرجات المعيارية أو المئين لدرجة الطالب على كل قدرة فرعية أو اختبار نفسي أو تحصيلي على هيئة نقاط، يصار إلى التوصيل بينها. ويبين الشكل التالي رقم (2) صفحة نفسية لطالب على أربع قدرات من بطارية الاستعداد الفارقة:



قدرة عددية قدرة مكانية اســتدلال

الشكل (2): صفحة نفسية لطالب على قدرات بطارية الاستعدادات الفارقة

وهناك عدد من الأشكال الذي تبنى به الصفحة النفسية، فقد يكون على هيئة أعمدة بيانية، يبين كل عمود فيها مستوى أداء الطالب على قدرة فرعية من القدرات التي تمثلها بطارية الإختبار، ويظهر عمـود يمثـل المئينـات أو الدرجات المعيارية على كل قدرة.

ملخص

أمكن تلخيص الأفكار التي قدمت في هذا الفصل في النقاط التالية:

- يتم اشتقاق المعايير بناءا على جماعات مرجعية تتسم بالحداثة والتمثيل والمواءمة.
- المعيار الوطني هو الأكثر شيوعاً واستخداماً لدى ناشري الإختبارات.
- يجب استخدام معايير متوسط المدرسة إذا اتجهت النية نحو مقارنة أداء إحدى مدارس المنطقة التعليمية بأداء مدارس مناطق تعليمية أخرى.
- تشير الرتبة المئينية للشخص إلى نسبة درجات أفراد المجموعة التي تقع تحت درجته. وتتصف بسهولة التفسير، إلا أنها تعاني من عيب الوحدات غير المتكافئة
- الدرجات المعيارية (Z) والدرجات التائية (T) هي درجات مشتقة بمتوسط مقداره صفراً و (50)، وبانحراف معياري مقداره (1) و (10) على التوالي.
- درجات معامل الذكاء الانحرافي هي درجات معيارية بمتوسط مقداره (100)، وبالاعتماد على الإختبار، بانحرافات معيارية تكون عادة (ولكن ليس دائماً) بمقدار (15) و (16).
- التساعيات المعيارية هي درجات مشتقة مسواة بمتوسط مقداره (5) وانحراف معياري مقداره (2).

- الفرق بين أي درجتين زائيتين في أي موقع في التوزيع يقابل فرقا ثابتا في الدرجات الخام.
- مجموع الدرجات الزائية يساوي صفرا، والإنحراف المعياري لها يساوي (1). بينما مجموع مربعات الدرجات الزائية يساوي عدد قيم عناصر العينة (n).
- تحوي الدرجات الزائية قيما موجبة وسالبة وكسورا. الأمر الذي لا يكون يسيرا على أولياء الأمور، أو العامة من الناس استخدامها أو تفسيرها.
- تعد الصفحة وسيلة مفيدة جداً في التفسير عندما يتوفر لدينا عدة درجات لشخص واحد. وعند استخدام الصفحة، يجب أن تحول جميع الدرجات إلى درجات معيارية من نفس النوع، وتبنى على مجموعة معيارية واحدة. إضافة إلى ذلك، يجب الإشارة إلى مقدار الخطأ في قياس كل قدرة.

أسئلة وتمارين

س1: هل يمكن للمعايير أن تكون حديثة وممثلة (معبرة) من غير أن تكون ذات علاقة؟ وضح إجابتك.

س2: هل تبدو الأرقام التالية معقولة؟ بين ذلك.

الرتبة الميئنية الوطنية للطالبة سميرة **95**

الرتبة المئينية المحلية للطالبة سميرة **85**

المتوسط المئيني المدرسي لمدرسة سميرة **35**

س3: ما هي المعلومات الإضافية التي تحتاجها لمقارنة درجتك في اختبار نصف السنة الدراسية (**43**). ودرجتك في الامتحان النهائي (**62**)؟

س4: إذا كان لديك توزيعاً اعتدالياً للدرجات الخام، فأي الدرجات التالية هي الدرجة الأبعد عن المتوسط:

أ) التائية = **65** ب) الزائية = 2.0 −

ج) المرتبة المئينية = **90** د) التساعية = 7

س5: إذا كان المتوسط الحسابي لمجموعة من الدرجات= (**100**) والانحراف المعياري لها=(**16**)، أحسب:

(أ) الدرجة الزائية والدرجة التائية المقابلة لكل علامة من الدرجات الخام التالية التي تنتمي لتلك المجموعة:

108	**100**	**104**	**120**
132	**124**	**92**	**84**

(ب) أوجد الدرجات الخام والدرجات التائية المقابلة للعلامات الزائية التالية التي تنتمي إلى المجموعة المذكورة:

-1.0	**0.0**	**1.5**	**0.25**
-1.25	**0.50**	**-2.0**	**2.00**

س6: أحسب الانحراف المعياري لمجتمع من البيانات متوسط علاماته (**115**)، إذا كانت فيه الدرجة (**145**) تقابل الدرجة الزائية (**1.5**) ؟

س7: في إمتحان للإحصاء، كان متوسط علامات الطلبة عليه (**75**) والانحراف المعياري لها (8) وكانت علامات كل من أحمد=(83)، ويوسف =(**75**)، وعلامة سميره الزائية=(**.75**) وعلامة عبدالله التائية=(90)، فكيف يكون ترتيب هؤلاء الطلبة تنازليا؟

س8: إذا كانت الدرجة الخام لمفحوص ما تساوي (**176**) ومتوسط مجموعته يساوي (163) والانحراف المعياري لها يساوي (26)، فكيف يمكن التعبير عن هذه الدرجة بدلالة:

أ) الدرجة الزائية

ب) الدرجة التائية

ت) الدرجة المئينية

س9: كيف يمكن ترتيب الدرجات التالية تصاعديا ؟

أ) $z = 1.5$

ب) $T = 30$

ج) p_{10}

س10: إذا كانت الدرجات التالية تائية (20 ، 30، 40 ، 50، 90، س)، فما مقدار الدرجة (س) ؟

س11: تأمل الدرجات الزائية التالية (صفر، -1 ، -.5 ، 2، -1.5 ، س) ، ثم أحسب مقدار الدرجة (س) ؟

الفصل الثامن

أدوات التقييم والقياس في العلوم السلوكية

Assessment and Measuremnt Tools in Behavioral Sciences

- مقدمة
- مقدمة
- أدوات التقييم والقياس في العلوم السلوكية
- تصنيف أدوات التقييم والقياس

أولا: حسب مجالات السلوك الإنساني

- أدوات التقييم والقياس في المجال المعرفي (الاختبارات)
- أدوات التقييم والقياس في المجال الانفعالي (الوجداني)

ثانيا: تصنيف أدوات التقييم والقياس حسب الصفة والهدف

- استخدامات أدوات التقييم والقياس
- ملخص
- أسئلة وتمارين

8

المخرجات التعليمية

بعد دراسة هذا الفصل يتوقع من الطالب أن يكون قادرا على أن

- يصنف أدوات التقييم والقياس حسب مجالات السلوك الإنساني.
- يقارن بين طرق قياس الاتجاهات.
- يقارن بين طرق قياس الشخصية.
- يقارن بين طرق قياس الميول.
- يقارن بين أنواع الأدوات الاسقاطية.
- يصنف أدوات القياس حسب الصفة والهدف.
- يصف دور أدوات التقييم والقياس في اتخاذ عدد من القرارات في الإختيار، والتصنيف، التسكين، التشخيص والعلاج، الإرشاد والتوجيه، وتطوير البرامج.
- يميز أوجه الشبه والاختلاف بين اختباري معياري المرجع ومحكي المرجع.
- يحدد نوع المعلومات المناسبة لاتخاذ قرار معين.

مقدمة

تعددت أدوات التقييم والقياس والتقويم وتنوعت حسب نوع المعلومات المطلوب الحصول عليها، وتبعا للهدف الذي يتوقع من القياس تحقيقه أو من التقويم بلوغه. فقد يكون الإختبار التحصيلي في مادة دراسية معينة هو موضع الاهتمام، وربما يكون المطلوب تقريرا ذاتيا يكتبه الطلبة أو المعلمون أو الأخصائيون عن أنفسهم أو عن غيرهم من المعنيين، وربما يكون ذلك بناءا على ملاحظات تم القيام بها أو مقابلة أجريت مع مجموعة من الأفراد موضع البحث والدراسة. وقد يكون المطلوب كذلك اختبار قدرات أو استعداد أو قائمة عبارات تقيس اتجاها ما نحو مسألة اجتماعية أو تربوية أو حالة نفسية شائعة. ومهما يكن من أمر، فإن أي أداة قياس يتم استخدامها تحتاج إلى مبرر عملي ومنطقي.

أدوات التقييم والقياس

اتضح مما سبق أن عملية القياس ترتبط عموما بجمع البيانات عن الفرد بهدف فهم وتفسير سلوكه ومن ثم تنمية أو تعديل ذلك السلوك (Anastasi, 1968). ومعروف كذلك أن السلوك الإنساني كل متكامل، يشكل منظومة متكاملة من الميادين المعرفية والإنفعالية والمهارية. الأمر الذي أدى إلى ضرورة تطوير أدوات قياس تتلاءم وتتوافق مع طبيعة وخصائص كل مجال من هذه المجالات. إضافة إلى أنه ربما يحتاج فهم السلوك الإنساني إلى معلومات كمية وبيانات نوعية.

تصنيف أدوات التقييم والقياس

في البداية، من الضروري التمييز بين مفهومي الإختبار Test والمقياس Scale. إذ أن الإختبار إجراء منظم لقياس عينة من السلوك، أو

هو إجراء منظم لملاحظة ووصف سمة أو أكثر من سمات الفرد بالاستعانة بمقياس أو نظام تصنيف معين. ويحدث أحيانا خلط بين مفهومي الإختبار والإمتحان Examination. إذ أن الإختبار أشمل من الإمتحان، ويتعلق الأخير بقياس الأداء في المجالات التحصيلية.

من ناحية أخرى، إذا كان الإختبار مرتبطا أكثر في قياس الأداء في المجالات المعرفية (الذكاء والإستعدادات والتحصيل)، فإن المقياس أداة قياس للجوانب المعرفية وكذلك الوجدانية أو تلك التي يتم قياسها على سلم يتألف من عدد من الدرجات (3)، أو (5)، أو (7). وإذا كان للإختبار نموذج تصحيح لفقراته، أي أن هناك إجابة نموذجية لفقرات الإختبار، فليس هذا ممكن للمقياس، فالاجابة على المقياس لا تستحمل الصح والخطأ، فهي تبين درجة انطباق العبارة وشدتها على المستجيب.

يمكن تصنيف أدوات القياس بأكثر من طريقة، فقد رأينا أن يتم ذلك حسب جوانب السلوك الإنساني (معرفي، وجداني (انفعالي)، مهاري) من ناحية ، وأن يصار إلى تصنيفها حسب عدد من العوامل من مثل: موضوع الأداة، وطريقة الإجابة المتوقع أن يعطيها الممتحن، أو حسب الجهة التي ستسجل أو تلاحظ السلوك موضع البحث، أو حسب طريقة تفسير النتائج. فالإختبارات مثلا تصلح في الغالب في قياس الجوانب المعرفية وهي الأكثر شيوعا وتطورا واستخداما، في حين هناك أساليب أخرى لا تقل أهمية عن الإختبارات كالمقاييس التي تستخدم في قياس الاتجاهات والشخصية وتقويم المهارات والميول وغيرها. وبشكل دقيق يمكن تصنيف أدوات التقييم عموما حسب مجالات السلوك الإنساني وحسب الصفة والهدف كما هو مبين في المخطط التالي:

مخطط يوضح تصنيف أدوات القياس النفسي في ضوء عدد من المتغيرات

مجالات السلوك الإنساني	المجال المعرفي	غير رسمية	اختبار من إعداد معلم المدرسة أو أستاذ الجامعة أو المدرب...		
			اختبار من إعداد الاخصائي النفسي أو الاجتماعي أو المرشد ...		
		رسمية	التحصيل	اختبارات تشخيصية	اختبار أيوا للمهارات الأساسية IBST
				بطاريات الإختبارات المسحية	بطارية كاليفورنيا للاختبارات CAT
				الإختبارات المسحية المتعلقة بمادة دراسية	
				اختبارات تحصيل لأغراض التنبؤ	
			القدرات	مقاييس لفظية أو غير لفظية	لفظي:كاليفورنيا،كالمان وأندرسون. غير لفظي: مصفوفات رافن المتتابعة، اختبار غودانف-هاريس لرسم الشخص، بيتا iii
				مقاييس فردية أو مقاييس جمعية	فردي: بينيه ، ويكسلر جمعي: بيتا، أوتيس- لينون
				مقاييس ذكاء عقلي وذكاء عاطفي	
			الإستعدادات	اختبارات استعداد أكاديمية	بطاريتا DAT و GRE
				اختبارات استعداد مهنية	اختبارات استعدادات منفصلة: فني، موسيقي، كتابي، يدوي، ميكانيكي
					بطاريات استعدادات: فارقة وعامة
					اختبارات موقفية: محاكاة ومواقف فعلية.
				اختبارات التفكير الابتكاري	بطارية اختبار تورنس للتفكير الابتكاري

				بطارية اختبارات الانتاج التباعدي/جيلفورد
				اختبار التداعي الطليق/مدنيك
	المجال الوجداني (الانفعال)	مقاييس الاتجاهات	أسلوب التقدير الجمعي (ليكرت)	
			أسلوب الفترات المتساوية ظاهريا (ثيرستون)	
			أسلوب تمايز معاني المفاهيم (أوزجود)	
			أسلوب التصنيف الترتيبي Q (ستيفنسون)	
			أسلوب البسط (كومبس)	
		قوائم الميول	أسلوب الاستناد لأسس نظرية	استبيان لي- ثروب للميول المهنية
			أسلوب الاستناد لفقرات متجانسة	قائمة كيودر للتفضيل المهني
			أسلوب الاستناد إلى التحقق الامبريقي	استبيان سترونج-كامبل للميول
		مقاييس العلاقات الاجتماعية	طريقة مورينو	مقياس العلاقات الزوجية والأسرية
				مقياس المسافة أو البعد الاجتماعي
				المصفوفة الاجتماعية
			طريقة جاردنر و تومسون	
		قوائم الشخصية	مقاييس محددة البناء وواضحة الهدف	ما يبنى على أسس نظرية
				ما يبنى على أسس أمبريقية
			مقاييس إسقاطية	أدوات تركز على المحتوى
				أدوات تركز على الشكل
				أدوات تتضمن المحتوى والشكل

			مقاييس غير محددة البناء لكنها واضحة الهدف	مقاييس أساليب الملاحظة
				مقاييس أساليب المقابلة
				مقاييس مفهوم الذات
			مقاييس محددة البناء لكنها غامضة الهدف	مقاييس الأنشطة الفسيولوجية
				مقاييس الإدراك
حسب الصفة والهدف	حسب تحديد المؤثر والاستجابة		(1) أدوات إسقاطية ،	(2) أدوات محددة البناء
	حسب مستوى الأداء		(1) أدوات ذات أقصى أداء ،	(2) أدوات الأداء العادي
	حسب عدد الأفراد		(1) أدوات فردية،	(2) أدوات جمعية
	حسب شكل الفقرات		(1) الإختبارات الإختيارية،	(2) اختبارات الإجابة المصوغة
	حسب الزمن المخصص للاستجابة		(1) اختبار سرعة	(2) اختبار قدرة
	حسب طريقة التصحيح		(1) موضوعية	(2) ذاتية
	حسب كيفية ظهور الأداء		(1) أدوات لفظية	(2) أدوات غير لفظية
	حسب الجهة التي تطور الأداة		(1) اختبارات مقننة أو منشورة	(2) اختبارات غير رسمية
	حسب تحديد المجال الذي يتم قياسه		(1) اختبار عينة	(2) اختبار مؤشر
	حسب السلوك المقاس والمستقبلي		(1) اختبار ممثل	(2) اختبار متنبىء
	حسب طريقة تفسير النتائج		(1) مقاييس معيارية المرجع	(2) مقاييس محكية المرجع

أولا : تصنيف أدوات التقييم والقياس حسب مجالات السلوك الإنساني

أدوات التقييم والقياس في المجال المعرفي (الإختبارات Tests) :

يعرف براون (Brown, 1976) الإختبار بأنه مجموعة الإجراءات المنظمة لقياس عينة من السلوك بهدف موازنة أداء الفرد بمعيار أو بمستوى أداء محدد. ويشتمل هذا التعريف على مجموعة العمليات والمفاهيم المرتبطة بالإختبارات من مثل التقنين، والموضوعية، وعينة السلوك، والمعيار والمحك. ومهما يكن من أمر فإنه يمكن تصنيف الإختبارات أساسا في نوعين رئيسين حسب الجهة التي تقوم على تطويرها أو إدارتها هما:

- **الإختبارات غير الرسمية** Informal أو ما تعرف بالإختبارات من إعداد الشخص المستخدم (معلم أو أخصائي) Man-made Tests. تلك الأدوات التي يشيع تطويرها واستخدامها من قبل معلمي المدارس وأساتذة

الجامعات، بقصد الاستخدام في تقويم أداء المتعلمين أو تشخيص مستوى تحصيلهم. وغالبا ما يصعب تعميمها أو الاعتماد عليها لأغراض تقويم نوعية برامج تعليمية أو لاتخاذ قرارات استراتيجية أكثر شمولية.

- **الإختبارات المقننة أو المنشورة** Standardized Tests والتي تعرف أحيانا بالإختبارات الرسمية Formal tests. وتكون تلك الإختبارات على هيئة اختبارات تحصيل أو قدرة أو استعداد:

أ. **اختبارات التحصيل Achievement tests** تهدف إلى قياس ما تعلمه المتعلم على مدى فترة زمنية معينة في موضوع معين أو عدة موضوعات، كامتحان الثانوية العامة في عدد من البلدان. كما أن هناك أنواع عديدة من اختبارات التحصيل المقننة من مثل:

i. **الإختبارات التشخيصية** Diagnostic tests: فقد تصمم الإختبارات المقننة لتشخيص صعوبات تعلم الطلبة في مادة دراسية معينة، بقصد تحديد جوانب الضعف ونواحي القوة في تحصيل الطالب. وغالبا ما تكون اختبارات تشخيص في المهارات الأساسية في القراءة أو الهجاء أو الحساب من أجل تقديم تعليم علاجي مناسب، وتصنف الإختبارات التشخيصية على أنها مرجية المحك Criterion-Referenced Tests. ومن أمثلة هذا النوع من الإختبارات اختبار "أيوا للمهارات الأساسية Iowa Test of Basic Skills".

بطاريات الإختبارات المسحية Survey tests: وتكون هذه البطاريات على هيئة مجموعة من الإختبارات المسحية التي يتعلق كل منها بمادة دراسية معينة، ومخصصة لمرحلة تعليمية معينة. وهناك بطارية الإختبارات المسحية للمرحلة الابتدائية (قراءة، حساب، علوم، دراسات اجتماعية، هجاء)، وأخرى للمرحلة الثانوية (علوم، رياضيات، مواد اجتماعية). وتكون هذه البطاريات من نوع اختبارات معيارية المرجع Norm-Referenced-Tests. ومن أمثلة هذه البطاريات العالمية: "بطارية كاليفورنيا للاختبارات المسحية California Achievement

Test (CAT) "، و "بطارية ستانفورد للمهارات الأكاديميـــة Stanford Test of Academic Skills (TASK) ". وجدير بالذكر أنه من الممكن استخدام هذه البطاريات في عمليات التشخيص.

ii. **الإختبارات المسحية المتعلقة بمادة دراسية واحدة:** وتهدف إلـــى قيـــاس التحصيل في مادة دراسية واحدة فقط. وغالبا ما تكون فقراتها أكثر عـــددا وأشمل توزيعا وأكثر تجانسا من الفقرات في البطاريات المـــسحية ســـالفة الذكر.

iii. **اختبارات تحصيل لأغراض التنبؤ:** وهذه تلعب دور اختبارات الاســـتعداد، إذ يستخدم مستوى الأداء عليها في التنبؤ بمستوى التحصيل اللاحق فـــي مادة الإختبار.

ب. **اختبارات القدرة** Ability Tests وأفضل مثال على اختبارات القـــدرة هي مقاييس الذكاء. ويمكن تصنيفها حسب البنية النظرية (الذكاء كمفهـــوم وصفي، كمجموعة من القدرات المترابطة، كعامل عـــام، كبنيـــة ثلاثيـــة الأبعاد، كنظام هرمي من العوامل) أو حسب عدد الـــذين يطبـــق علـــيهم الإختبار، فهناك اختبارات فردية من مثل مقياس ستانفورد بينيه للـــذكاء ومقياس ويكسلر لذكاء الراشدين (WISC)، أو مقيـــاس ويكـــسلر لـــذكاء أطفال ما قبل المدرسة (WPSI). وهناك اختبارات جمعية لفظيـــة مثـــل اختبار أوتيس لينون للقدرة العقلية العامة، ومقاييس جمعية غير لفظية مثل اختبار المصفوفات المتتابعة لـــرافنRaven Progressive Matrices ، واختبار بيتا III للذكاء غير اللفظي، واختبار غود إنف-هـــاريس للرســـم .Good Enough-Harris Drawing Test

وقد برز حديثا مفهوم الذكاء العاطفي أو الإنفعالي Emotional Intelligence مؤخرا ليكون بديلا أو رديفا للذكاء العقلي في اتخاذ قرارات تتعلق في التنبـــؤ بالنجاح في أداء مهمات ومسؤوليات معينة. فقد عرف الذكاء العاطفي بأنه قدرة

الشخص على فهم وضبط وتوجيه عواطفه وعواطف الآخرين (Goleman, 1995).

ت. **اختبارات الاستعداد** Aptitude tests التي تهدف إلى قياس مدى قدرة الشخص على القيام بعمل مستقبلي، كشغل وظيفة ما أو ممارسة مهنة معينة أو الالتحاق ببرنامج تدريبي أو إرشادي معين. ومن أمثلة اختبارات الاستعداد: بطارية اختبارات الإستعدادات الفارقة Differential Aptitude Test (DAT) وهي من أكثر بطاريات الإستعدادات شهرة واستخداما، وتشمل ثمانية اختبارات فرعية مثل الاستدلال اللفظي، والقدرة العددية، والاستدلال المجرد، والسرعة، والاستدلال الميكانيكي وغيرها. وبطاريات اختبارات سجل الدراسات العليا GRE التي تتألف من ثلاثة اختبارات فرعية (القدرة اللغوية، القدرة العددية، القدرة التحليلية) وتستخدم بشكل واسع في عمليات قبول الطلبة في برامج الدراسات العليا في الجامعات الأمريكية وبعض الجامعات في أنحاء مختلفة من العالم. كما أن اختبارات التفكير الابتكاري Creativity التي تندرج تحت اختبارات الاستعداد ومن أمثلتها: بطارية اختبارات تورانس للتفكير الابتكاري Torrance Tests of Creative Thinking وهي الأكثر ملائمة للواقع التربوي، وتصلح للتطبيق على الأفراد من مستوى الحضانة إلى المراحل الدراسية العليا. وتتألف القدرة على التفكير الابتكاري من ثلاث قدرات فرعية هي الطلاقة والأصالة والمرونة. وهناك حقيبة اختبار فرانك وليامز التي تهدف إلى تقييم القدرة على التفكير التباعدي وكذلك الشعور التباعدي.

والمعروف أن الإختبارات المقننة يتم إعدادها من قبل فريق من المختصين ولصالح مؤسسة بحثية أو أكاديمية أو حكومية معينة يتم بعد ذلك نشرها. وتخضع الإختبارات المقننة إلى معايير مشددة قبل نشرها، ولها تعليمات محددة لأغراض التطبيق والتصحيح وتفسير نتائج الأداء عليها. وجدير بالذكر أنه يتم مراجعة الإختبارات وتعديلها بشكل دوري لمراعاة المسائل

الثقافية ومدى تأثيرها في أفراد المجتمع. ولاشك أن تقنين مقياس ما، يتطلب ترجمته إلى لغة المجتمع الذي ستستخدم من أجله، ثم يجري مراجعة وتدقيق وتعديل مفردات الإختبار لتتلاءم مع فهم وثقافة أفراد ذلك المجتمع. كما يتضمن الإختبار المقنن دليل استخدام (manual) يصف الخلفية النظرية للاختبار وتركيبه، ووصف الأبعاد التي يتكون منها ذلك الإختبار، هذا بالإضافة إلى تعليمات التطبيق والتصحيح واستعراض معايير تفسير مستوى أداء الممتحنين حسب متغيرات متنوعة كالجنس والفئة العمرية وغيرها.

أدوات التقييم والقياس في المجال الإنفعالي (مقاييس و قوائم)

تنوعت وتعددت كثيرا أدوات القياس في المجال الإنفعالي نظرا لتعدد وتنوع السمات النفسية التي يراد قياسها، ومن أمثلة تلك المقاييس تلك المتعلقة بالاتجاهات والميول والقيم والشخصية وغيرها. ومعظم هذه الأدوات تعرف على أنها أدوات تقرير ذاتي self-report tools ، بمعنى أن الشخص يستجيب على عدد من الأسئلة ليعبر عن نفسه (Gay, Mills, and Airasian, 2008). وبشكل عام، فإن هذه الأدوات بحاجة إلى تجريب وتطوير قبيل تطبيقها ميدانيا كي لا تكون أدوات مضللة. إذ يجب استخدامها بحذر شديد إذا لم يتضح أنها طورت حسب الأصول.

- **مقاييس الاتجاهات Attitudes Scales**

تعمل مقاييس الاتجاه على تحديد ما يعتقده أو يشعره أو يدركه الفرد عن نفسه، كما تحاول قياس الاتجاهات نحو الذات أو نحو الآخرين أو نحو أنشطة معينة أو تجاه واقع معين. وهناك عدد من أساليب بناء مقاييس الاتجاهات مثل:

1. أسلوب التقدير الجمعي مثل سلم ليكرت Likert Scale
2. أسلوب الفترات المتساوية ظاهريا أو ما يعرف بسلم ثيرستون Thurston Scale

3. سلم تمايز التباين اللفظي (السيمانتي الفارق) Semantic Differential Scales وقام بتطويره أوزجود Osgood
4. الأسلوب التراكمي أو ما يعرف بسلم جتمان Guttman Scale
5. أسلوب التصنيف الترتيبي (Q) لستيفنسون Stephenson
6. أسلوب البسط **Unfolding Techniques** واقترحه كومبس **Coombs**.

وبالرغم من شيوع التعامل مع فقرات مقاييس الاتجاهات مجتمعة، إلا أنه من الممكن التعامل مع كل عبارة على حدة، وهذا مهم في حالة أن تكون طبيعة الاتجاه موضع البحث لا تتطلب توفر كافة العناصر المكونة للاتجاه لدى المستجيب. وبشكل عام يشيع استخدام سلالم التقدير في قياس درجات امتلاك الأشخاص لسمات نفسية مثل القلق أو التوافق أو التكيف أو مفهوم الذات وغيرها من ناحية ، وفي عمليات التشخيص من ناحية أخرى. كما يغلب استخدام هذه السلالم من فئة الخمس درجات، مشيرا إلى أنه من الممكن تحويل سلالم التقدير إلى قوائم فحص Checklists والعكس صحيح.

ففي أسلوب التقدير الجمعي أو ما يعرف بسلم **ليكرت**، يطلب من الفرد أن يستجيب لسلسلة من العبارات لبيان ما كان يوافق عليه المستجيب بشدة (5 درجات)، أو يوافق (4 درجات)، أم غير متأكد (3 درجات)، أم لا يوافق (درجتان)، أو لا يوافق بشدة (درجة واحدة) على كل عبارة.

وجدير بالذكر أن جميع الفقرات في سلم ليكرت لها نفس الوزن أو نفس الأهمية. ويكون تقدير درجة الفرد على مقياس الاتجاه بطريقة ليكرت بإيجاد مجموع هذه القيم أو الدرجات علي جميع العبارات. ويراعى عند جمع هذه القيم الانتباه إلي العبارات السالبة التي يجب عكسها أولا. كما أن أقصى درجة يتوقع أن يحصل عليها الفرد تساوي حاصل ضرب عدد الفقرات بأقصى درجة على درجات السلم (وهي 5 في المثال أعلاه)، وقد تكون (3) أو (7) في بعض الأحيان. ويمكن أن تكون مستويات مقياس ليكرت في التعبير عن عبارات مقياس الاتجاه بأشكال متنوعة مثل:

ضعيف	**مقبول**	**جيد**	**جيد جدا**	**ممتاز**
مطلقا	**نادرا**	**أحيانا**	**غالبا**	**دائما**
1	**2**	**3**	**4**	**5**
منخفض جدا	**منخفض**	**متوسط**	**عالي**	**عالي جدا**
موافق		**بين بين**	**غير موافق**	

وبالنسبة لأسلوب الفترات المتساوية أو ما يعرف بسلم **ثيرستون**، تعطى أهمية معينة لكل فقرة حسب ما تستحقه، ويتم التمييز بين الفقرة والأخرى بواسطة قيم عددية يتم تحديدها بطرق خاصة، ويصار غالبا عند تطوير هذه المقاييس إلى إبقاء الفقرات الرئيسة واستبعاد الفقرات قليلة الأهمية. ويمكن أن يتحقق ذلك في قياس بعض سمات الشخصية وقياس الاتجاهات، ويشذ عن ذلك قياس المهارات. إذ لا يمكن في هذه الحالة إهمال فقرات والتركيز على البعض الآخر، لكونها (المهارة) تشكل بمجملها نظاما متكاملا.

وبشكل عام تعرض الفقرات المراد تطويرها في مقياس على مجموعة من المحكمين ويطلب منهم تصنيف كل عبارة حسب درجة قبولها على سلم يتألف من 11 فئة يتراوح بين أدنى درجة من الرضا (1) إلى أقصى درجة رضا (11). وعليه يتم اختيار الفقرات التي استحقت (7) درجات أو أكثر وتعتبر بعد ذلك على نفس الأهمية. أما إذا كانت الفقرات محددة مسبقا كما في المهارات، فيمكن أن تأخذ كل فقرة وزنا خاصا حسب قيم وسيط كل منها.

أما السلم **السيمانتي الفارق (التمايز اللغوي)**، يطلب من الفرد أن يحدد تقديرا كميا للموضوع أو المفهوم الذي يقيسه مقياس الاتجاه وذلك باستخدام أزواج من الصفات المتضادة ثنائية القطب. وقد استنتج أوزجود من خلال التحليل العاملي على عدد من المفاهيم ثلاثة عوامل أساسية تنطوي عليها تلك التقديرات هي:

التقييم مثل: (حسن – رديء)، (عادل – غير عادل)، (ناجح– راسب).

القوة مثل : قوي–ضعيف)، (متشدد– متساهل)، (كبير–صغير) .

النشاط مثل: (سريع–بطيء)، (ساخن–بارد)، (متوتر–مسترخي).

ويحدد المستجيب استجابته بوضع نقطة على المتصل المستمر الذي يمتد بين طرفي المتصل التي تتراوح بين (1، 7) أو بين (1 و 9)، وهذه النقطة تمثل اتجاهه. فقد يستخدم مقياس الاتجاه نحو نظام الساعات المعتمدة في جامعة ما الصفات الآتية:

فعال ------------------ **غير فعال**

عادل ------------------ **غير عادل**

أو قد نستخدم مقياس الاتجاه نحو قراءة القصص البوليسية من خلال صفات متضادة ثنائية القطب مثل:

مسلية - - - - - - - - - - - **مملة**

مثيرة - - - - - - - - - - - - - **مثبطة**

وهناك مثال آخر يتعلق بقياس الاتجاه نحو العولمة لدى طلبة الجامعة من خلال الصفات القطبية مثل:

استقلال - - - - - - - - - - - - - **عبودية**

عدل - - - - - - - - - - - - - - **ظلم**

ديمقراطية - - - - - - - - - - - - - **دكتاتورية**

وفي الممارسة، كثيرا ما تتجاهل هذه الأبعاد، أو يحل محلها أبعاد أخرى يرى أنها أكثر ملائمة لموقف معين دون آخر. ولكل موضع على المتصل المستمر قيمة أو درجة. وبإيجاد مجموع هذه القيم على جميع البنود يمكن تحديد ما إذا كان اتجاه المجيب موجبا أم سالبا. ويكون للمقاييس السيمانتية الفارقة عادة خمس مسافات أو سبع مع وجود الإتجاه الوسط الذي يعطى درجة صفر كما في الشكل التالي:

نظام الساعات المعتمدة

فعال	**(3)**	**(2)**	**(1)**	**صفر**	**(-1)**	**(-2)**	**(-3)**	**غير فعال**
عادل	**(3)**	**(2)**	**(1)**	**صفر**	**(-1)**	**(-2)**	**(-3)**	**غير عادل**

فإذا وضع المستجيب على البندين إشارة عند الرقم (3)، فهذا يدل على اتجاه موجب جدا نحو النظام القضائي. وفي حالة اختياره الرقم (صفر) مثلا، فهذا يعني أنه لا يملك موقفا محددا نحو القضاء، أو أنه غير معني به.

ويستخدم أسلوب **التصنيف الترتيبي Q-Sort لستيفنسون** في أغراض الإرشاد والعلاج النفسي والتغيرات السلوكية، والمقارنة بين تفضيلات أو آراء مجموعة من الأفراد في موضوع ما. ويتلخص هذا الأسلوب باعطاء الفرد مجموعة من البطاقات ويطلب منه ترتيبها في مجموعات وفقا لموقعها النسبي على متصل يتراوح بين "الأكثر أهمية" و "الأقل أهمية" أو " الأفضل" و "الأسوأ". وهناك حد أعلى لعدد البطاقات المسموح بها في كل مجموعة (علام، **2000**).

وبالنسبة لسلم **جتمان**، فهو قليل الاستخدام في تصميم أدوات التقويم بالرغم من أنه يراعي مسألة هامة في القياس لا تتوفر في غيره من مقاييس الاتجاهات الأخرى. إذ أنه لا يمكن تحديد الفقرات التي أجاب عنها الفرد إجابة صحيحة في المقاييس الأخرى إذا عرفنا علامته الكلية ذلك لأنها غير متسلسلة منطقيا.

وبالرغم من صعوبة استخدام هذا الأسلوب في كثير من السمات والقدرات والمهارات، لكنه ممكن أحيانا، فقد يتم ترتيب الفقرات وفق تسلسل منطقي يعرف من خلاله موقع كل فقرة بالنسبة للفقرات الأخرى، كما في المهارات الصغيرة أو المتسلسلة التي تشكل بمجموعها مهارة

معقدة. فقد تكون المهارة تراكمية أو متطلبا سابقا لإتقان مهارة لاحقة. فإذا قفز شخص حاجزا ارتفاعه مترا واحدا فهو بالتأكيد قادر على قفز الحواجز التي يقل ارتفاعها عن المتر. كما أن عدم رغبة الطالب لرؤية المعلم تأتي بعد عدم رغبته في حضور محاضرته، وهذه تأتي بعد عدم رغبته في الحديث معه وهكذا (عودة، **2000**).

كما يحاول هذا السلم أن يحدد ما إذا كان الاتجاه أحادي البعد، أي إذا أسفر عن مقياس تراكمي. فالفرد الذي يجيب على فقرة معينة موافقا، فإنه سيوافق على كافة الفقرات التي تتصل بها وتسبقها. فلو وافق أحدهم على العبارة رقم (3) فهذا يعني أنه يوافق ضمنا على العبارتين (2) و (1) ولا يوافق على ما بعدها (**Gay, Mills, and Airasian, 2010**).

وبين عامي 1953 و 1964 اقترح كومبس **Coombs** أسلوبا جديدا في إطار نظريته التي أطلق عليها نظرية البيانات، يستخدم في استخلاص سمة كامنة مثل الاتجاهات التي تنطوي عليها تفضيلات مجموعة من الأفراد. ويطلب من الأفراد اختيار الفقرات التي يرون أنها أكثر تفضيلا وميلا. ومن مقاييس الاتجاهات المعروفة والمنشورة باللغة العربية: مقياس الاتجاهات الوالدية ومقياس الاتجاهات العائلية ومقياس الاتجاه نحو العلم والمهن العلمية.

والجدير بالذكر أن هناك أدوات أخرى لقياس الاتجاهات من مثل أساليب التصنيف الترتيبي، وأسلوب البسط، والأساليب متعددة الأبعاد، وأسلوب اختيار الخطأ، وأسلوب القبول الظاهري، وأساليب استجابات الآخرين، إضافة إلى أساليب تعتمد على اختبارات الإدراك والذاكرة.

- **قوائم الميول Interests Inventories**

هي أدوات قياس يطلب فيها من الفرد أن يبين ما يميل إليه من أنشطة وما لا يميل إليه. ويمكن تصنيف الميول في أربع مجموعات استنادا إلى الأسلوب الذي يتبع في الحصول على المعلومات على النحو التالي:

- i. **ميول يعبر عنها الفرد Expressed Interests**
- ii. **ميول تنعكس في سلوك الفرد Manifested Interests**
- iii. **ميول يستدل عليها من خلال مستوى الأداء على اختبارات معينة كالتحصيل Inferred Interests**
- iv. **ميول تقاس بالاستبانات المقننة Inventorized Interests**

وتختلف أساليب بناء استبانات الميول وتقدير الدرجات عليها عن أساليب قياس المجالات المعرفية، حيث أن للأخيرة إجابة صحيحة واحدة مقبولة، بينما ليس كذلك بالنسبة لقياس المجالات الوجدانية. وهناك عدد من أساليب بناء استبانات الميول، مثل:

- **أسلوب الإستناد إلى أسس نظرية أو منطقية:** إذ يتم انتقاء فقرات أو مفردات الاستبانة اعتمادا على التحليل النظري للسمة أو الخاصية المراد قياسها من ناحية وكذلك أن تكون الفقرات سهلة التقدير . فمثلا الفقرة (أميل إلى قراءة المقالات الاجتماعية) أسهل تقديرا من الفقرة (أفضل المشاركة في الحفلات الاجتماعية مع الأصدقاء، عن المشاركة في الحفلات الكبرى). وقد جاءت استبانة الميول التي بناها لي ثروب **Lee Thrope** اعتمادا على هذا الأسلوب.
- **أسلوب الإستناد إلى بناء فقرات متجانسة:** فبعد أن يتم انتقاء الفقرات كما في الطريقة السابقة، يمكن التحقق من الاتساق الداخلي أو تجانس الفقرات. كما يمكن إخضاع الفقرات للتحليل العاملي للتوصل إلى عدد أقل من العوامل، والتوصل إلى أن الفقرات التي تشكل عاملا واحدا، تفسر ارتباطات عالية نسبيا بين الفقرات، وتشير إلى ميل معين. وقد جاء مقياس جيلفورد اعتمادا على هذا الأسلوب.
- **أسلوب الإستناد إلى التحقق الامبريقي:** ويتم انتقاء الفقرات استبيان ما على درجة تمييزها بين مجموعة من الأشخاص تنتمي

إلى مهنة معينة ومجموعة من عامة الناس أو مجموعات من مهن مختلفة. ويتطلب هذا الأسلوب تجميع عدد كبير نسبيا من الفقرات وتجريبها ميدانيا على مجموعة من الناجحين في مهنة معينة ومجموعة أخرى من عامة الناس. ومن مزايا هذا الأسلوب انه لا يمكن المستجيبين من تزوير إجاباتهم. ومن أشهر مقاييس الميول التي اعتمدت هذا الأسلوب هو قائمة سترونج-كامبل للميول.

وتفيد قوائم الميول عموما في توجيه الفرد وإرشاده إلى اختيار المجـال أو المجالات التي يحتمل بدرجة أكبر أن ينجح في ممارستها، وأن يرضى عن هذا العمل. ومن أكثر القوائم شيوعا:

- **قائمة سترونج-كامبل للميول** The Strong - Campbell Interest Inventory (SCII) وهو من أكثر القوائم شيوعا واستخداما. إذ تم إجراء العديد من البحوث والدراسات عليه **(Buros, 2001)**. ويعد سترونج بحق رائدا لحركة قياس الميول، كما كان بينيه رائدا لحركة قياس الذكاء. وقد وضع سترونج مقياسه الذي يتألف من 400 سؤال تدور حول نشاطات معينة، ويختار المستجيب واحدا من 3 بدائل هي: **(أحب)** أو **(لا أدري)** أو **(لا أحب)**. ومن أمثلة العبارات التي تتألف منها القائمة:

مقابلة الأشخاص لشغل وظيفة:	**أحب**	**لا أدري**	**لا أحب**
شراء السلع :	**أحب**	**لا أدري**	**لا أحب**

وتصحح الإجابات على هذا المقياس وفق (47) مهنة للرجال و(28) مهنة للنساء. وهناك مفتاح تصحيح لكل مهنة. وقد تم إعداد مفتاح التصحيح اعتمادا على الفرق بين إجابة أشخاص ناجحين في مهنة معينة وبين إجابة عينة من الناس ممن لا يعملون في تلك المهنة. وقد تطلب تطوير هذه المفاتيح إجراء عدد من البحوث عليها.

- **قائمة كيودر للتفضيل المهنـــي Kuder Preference Inventory** التي عربها أحمد زكي صالح وتقـيس الميـل نحـو العمـل الميكانيكي، والحسابي، والعلمي، والفني، والأدبي والموسـيقي ،والميل نحو العمل الكتابي والميل للخدمة الاجتماعية. تتـألف القائمة من مجموعات ثلاثية في كل منها 3 عبارات، يطلب من المستجيب أن يؤشر في كل مجموعة ثلاثية على العبارة التـي تشير إلى نشاط يجب أن يقوم به أكثر من غيره، والعبارة التي يجب أن يقوم بها أقل من غيرها. ومن أمثلة العبـارات التـي تتألف منها القائمة:

1. الذهاب إلى حفلة سينمائية

2. السير لمسافات طويلة في الغابة

3. الذهاب إلى معرض للمخترعات الحديثة

أو:

1. تركيب ساعة محطمة

2. الاحتفاظ بمجموعة من التقارير

3. رسم صورة

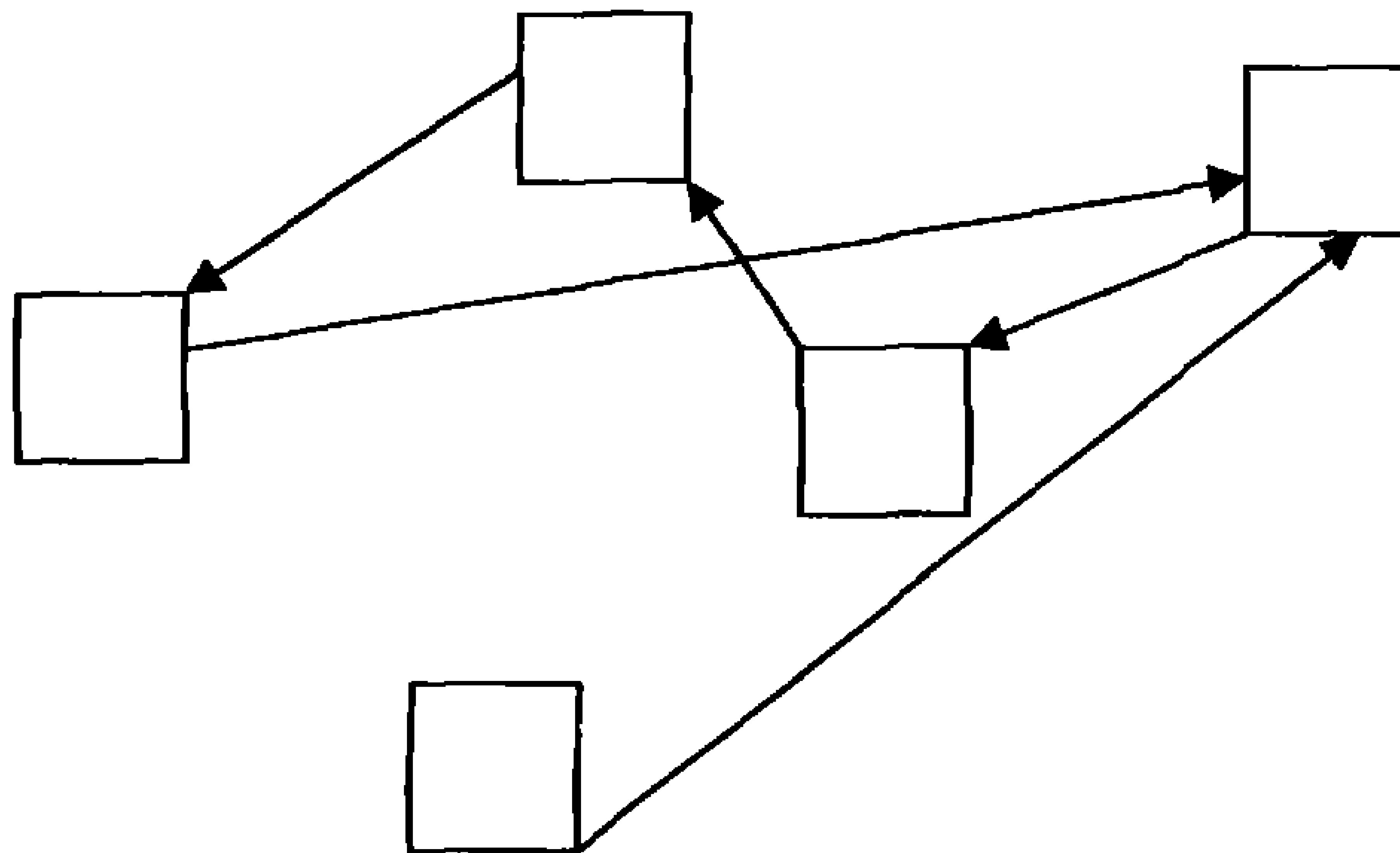
ويعطينا هذا المقياس أو القائمـة درجـات للميـول التاليـة: الترويحيـة، الميكانيكية، الحسابية، العلمية، الفنية، الموسيقية، الخدميـة، الاجتماعيـة، الإقناعية، والأدبية.

وجدير بالذكر أن مفاتيح التصحيح قد وضعت بناءا على العلاقات الداخلية بين العبارات. وقد بينت دراسة الإجابات عـن العبـارات أن عـددا مـن العبارات المتعلقة بالنشاط الإقناعي مثلا تنزع إلى أن تتجمع مع بعض. فإذا اختار أحدهم عبارة ما، فإنه يميل بدرجة كبيرة إلى اختيار بقية العبـارات الأخرى في المجموعة، والعكس صحيحا. كما ظهر أن ارتباط فقرات هذه المجموعة ببقية الفقرات الأخرى في المقياس ضعيف.

قياس العلاقات الاجتماعية Sociometrics

يهدف هذا النوع من المقاييس إلى دراسة التفاعل والتواصل الاجتماعي بين أفراد مجموعة، وقياس أنماط هذا التفاعل. وهو تعبير كمي للعلاقات الاجتماعية، يتمثل كتحويل العلاقات الوصفية بين الأفراد إلى أرقام وكميات. إنه مجال أساسي من مجالات علم النفس الاجتماعي الذي يعنى بدراسة البناء الاجتماعي للجماعات المختلفة. كما يهدف قياس العلاقات الاجتماعية إلى تقييم أنماط العلاقة التي تحكم تعامل الأفراد في مجموعة معينة، أو تجمع سكني، أو مهني أو تعلمي. ويقود الاسترشاد بنتائج هذه المقاييس إلى تكوين مجموعات يتوقع أن يكون أداؤها أو إنجازها أفضل. كما تستخدم هذه المقاييس في عمليات تأسيس فرق العمل الجماعي والأنشطة المدرسية والرياضية والاجتماعية والرحلات. ومن طرق قياس العلاقات الاجتماعية:

طريقة مورينو: وتقوم على مبدأ الإختبار السوسيومتري يجب أن يتألف من عدة أسئلة تطلب من المفحوص أن يحدد اختياره أو رفضه لبعض أعضاء الجماعة. ذلك بهدف تحديد مظاهر الألفة والحب والكره التي تحدد العلاقة القائمة بين كل فردين من أفراد الجماعة التي ينتمي إليها الفرد. فقد يتم توجيه أسئلة لكل فرد عن مع من تحب أنة تسافر؟ مع من تريد تناول طعام الغداء؟ شريطة أن تتوفر في الأسئلة عوامل الوضوح والشمول لمواقف متعددة وملائمة لمستوى الأفراد (الكندري، 1995).



وقد انبثق عن هذه الطريقة تمثيلا بيانيا أو شكليا للعلاقات الاجتماعية Sociogram يستخدم في تحليل بنية العلاقات بين أفراد المجتمع. ففي الرسم التالي يمكن ملاحظة أن الشخص (أ) اجتماعيا أو يدعى نجم المجموعة، لأنه الأكثر من بين الأفراد الآخرين ممن له علاقات مع جميع الأشخاص (ب، ص، ج) عدا الشخص (س). وأقلهم في ذلك الشخص (ص) ، ويمكن للشخص (أ) أن يتواصل مع الشخص (س) مثلا من خلال وساطة الشخص (ب) أو الشخص (ج) وهكذا:

وهناك عدد من شبكات العلاقات الاجتماعية حسب عدد الأشخاص من ناحية ونوع العلاقات من ناحية أخرى.

طريقة جاردنر وتومسون: وتستند على مبدأ أن الطريقة المثلى لقياس سوسيومترية الجماعات الصغيرة هي استخدام جماعة مرجعية كبيرة لصناعة المقياس السوسيومتري الذي يتم على أساسه الإختيار في الجماعات الصغيرة. إذ يطلب من تحديد اسم الشخص الذي قابله في حياته في مجموعته الصغيرة ويرغب أن يرافقه في رحلة علمية ويكتب اسمه على أقصى اليمين من خط مستقيم يمثل المقياس المفروض عمله وليكن (أ). ثم يحدد اسم الشخص الذي لا يرغب أن يرافقه في الرحلة ممن قابلهم في حياته وليكن (ب) ويكتب اسمه عند أقصى اليسار من نفس الخط المستقيم. ثم يطلب من الفرد أن يحدد اسم الشخص الذي يقع بين (أ) و (ب) وليكن (ج) وكذلك بين الشخصين (أ) و (ج)، وهكذا كما في الشكل التالي:

(أ) ــــــــــ **(ج)** ــــــــــ **(ب)**

كذلك يطلب من المفحوص أن يضع أسماء مجموعته الصغيرة في الأماكن المناسبة على هذا المقياس. بهذا تكون الدرجات المشتقة من طريقة جاردنر وتومسون أكثر دقة من تلك المشتقة من مقياس مورينو لإمكانية استخدامها بين الجماعات الصغيرة.

قوائم الشخصية Personality Inventories

تتألف قوائم الشخصية من مجموعة من الأسئلة أو العبارات التي تصف خصائص الشخصية وسماتها، ويطلب من الشخص أن يجيب عنها بـ (نعم) أو بـ (لا)، أو (أي تدريج آخر) لبيان ما إذا كانت العبارة تنطبق عليه أم لا. وقد تعددت أساليب قياس الشخصية وتعقدت كما في الأمثلة التالية:

- **مقاييس شخصية واضحة الهدف وذات بناء محدد**، وتقع في عدة أساليب منها تلك التي تبنى على أسس نظرية تتعلق بمحتوى السمة المقاسة. وخير مثال على هذا النوع من المقاييس هي قائمة وودورث للبيانات الشخصية Woodworth Personal Data Sheet (WPDS) وهي بمثابة تقرير ذاتي للشخصية، تم بناؤها أثناء الحرب العالمية الأولى، وكانت تهدف إلى تحديد الأفراد المجندين الذين يعانون من مشكلات انفعالية تعوق أدائهم للمهام العسكرية، وكانت تلك القائمة بديلا للمقابلة الشخصية الفردية، فهي أسلوب مقابلة جماعي باستخدام الورقة-القلم، وتتألف القائمة من 116 سؤالا يستجيب عنها الفرد إما بـ (نعم) أو بـ (لا) مثل:

هل تشعر دائما بأن صحتك جيدة ؟

هل تعاني من أحلام اليقظة ؟

هل منظر الدم يشعرك بالدوار ؟

هل تتلعثم أثناء المحادثة ؟

وهناك مقاييس تستند على أساس أمبريقي ومن أمثلتها قائمة مينيسوتا للشخصية متعددة الأوجه Minnesota Multiphasic Personality Inventory (MMPI) وهي من القوائم المشهورة في العالم حتى أنه كتب عنها أكثر من

700 بحث ودراسة (Buros, 2001) . وقد تألفت من 550 فقرة ، وتستخدم القائمة في التمييز بين الأفراد المكتئبين والفصاميين من جهة، والعاديين من جهة أخرى. ويستخدم هذا المقياس مع الكبار والمراهقين من سن 18 فأكثر. وهناك المقياس المعروف MMPI-A للمراهقين في الأعمار تحت 18 سنة. ويتألف المقياس عموما من ثلاثة مقاييس فرعية:

- **مقاييس لأغراض الصدق:** وتتألف من أربعة مقاييس هي: لا أستطيع أن أقول Cannot Say وتتألف من 30 عبارة أو أكثر والتي إذا لم يجب المفحوص عليها تعتبر النتائج مشكوك بها وربما غير صادقة. **ومقياس الكذب** L Scale وتتكون من 15 بند ويعالج الأخطاء البسيطة ونقاط الضعف التي يستطيع معظم الأفراد الاعتراف بها مثل: لا أحب كل الأفراد الذين أعرفهم، أو أقرأ كل عنوان في الجريدة اليومية. **ومقياس** F Scale ويتكون من 60 عبارة تستخدم كمؤشر لاتجاهات المفحوص نحو الإختبار والتعرف على مدى اتساق إجابة المفحوص مع إجابة معظم الأفراد. وهذا يقيس البارانويا والاتجاهات المضادة للمجتمع. **ومقياس** K scale ويتكون من 30 عبارة، تعتبر مؤشرا على محاولات المفحوص إنكار الأمراض النفسية وإظهار نفسه بصورة مقبولة أو المبالغة في الظهور بمظهر جيد.
- **مقاييس اكلينيكية** Clinical Scales عشر، هي: (1) توهم المرض (HS) Hypochondriasis ، (2) الاكتئاب Depression (D)، (3) الهستيريا Hysteria (Hy)، (4) الانحراف السيكوباثي Psychopathetic Deviate (Pd)، (5) الذكورة والأنوثة (Mf) Masculinity/Femininity ، (6) البارانويا Paranoia (Pa) ، (7) السيكاثينيا Psychasthenia (Pt)، (8) الفصام Schizophrenia (Sc)،

(9) الهوس الخفيف Hypomania (Ma)، (10) الإنطواء الإجتماعي Social Introversion (Si).

- **مقاييس المحتوى** Content Scales وهي (1) مقاييس القلق، (2) مقاييس المخاوف، (3) مقاييس الأفكار غير الواقعية، (4) مقاييس الأفكار القهرية، (5) مقاييس الغضب، (6) مقاييس مفهوم الذات، (7) مقاييس عدم الارتياح الاجتماعي، (8) مقاييس الكبت، (9) مقاييس المسؤولية الاجتماعية، (10) مقياس عدم التوافق.

وهناك عدد من مقاييس الشخصية المنشورة بالعربية مثل: قائمة آيزنك للشخصية و مقياس الشخصية للمرحلة الثانوية (اختبار كاليفورنيا).

مقاييس شخصية (الأدوات الإسقاطية Projective tools)

وتشكل هذه الأساليب اتجاها آخر في تقييم الشخصية يختلف عن الاتجاه المتعلق بقوائم الشخصية كما مر في البند السابق. وتتضمن الأدوات الاسقاطية عبارات وفقرات وأسئلة مصاغة لفظيا أو على هيئة رسومات أو أشكال أو مواقف غير محددة تماما من حيث المطلوب، وتستجر هذه الفقرات إجابات قد لا تظهر بنفس الاتجاه لو عرضت بشكل أسئلة صريحة. ويمكن تصنيفها كما يلي:

- **أدوات اسقاطية تركز على المحتوى** مثل اختبار تفهم الموضوع Thematic Apperception Test (TAT) الذي أعده مراي (Murray, 1938)، ويتم فيه عرض سلسلة من الصور تتميز بالغموض النسبي على الأفراد ، ثم يطلب من كل منهم أن يعبر عن ذلك في قصص يكتبها ومن ثم يصار إلى تحليلها. وتبرز أهمية هذا المقياس من حيث قدرته على الكشف عن الجوانب المخفية والعميقة من الشخصية حيث أنه يقلل من قدرة المستجيب على الكذب، لأن الهدف من المقياس غير واضح

للمفحوص. وكذلك فإن المقياس لا ينتمي إلى مدرسة واحدة من مدارس علم النفس. هذا بالرغم من أن ارتفاع تكلفة التدريب على التصحيح وطول فترة التطبيق كان أهم سلبية توجه ضد ذلك المقياس. ومن أمثلة البطاقات التي يتكون منها المقياس:

- البطاقة الأولى 1: تصف العلاقة بين الفرد ووالديه ونظرته لهم (مسيطرين، مساعدين، متفهمين، معتدلين في التربية). وتعطي البطاقة معلومات عن الحاجة للإنجاز وكيفية تحقيق ذلك والاستقلالية عن الأسرة أو الخضوع لها.
  - البطاقة التاسعة 9GF: وتعبر عن علاقة الفتاة بزميلاتها وتعالج مواضيع مثل حل الصراعات والخلافات، والغيرة، والتنافس بين الأخوات. كما قد تعبر عن الشكوك وعدم الثقة وقد تعبر عن أفكار البارانويا.
  - البطاقة العاشرة 10: إدراك المستجيب للعلاقة بين الرجل والمرأة وخاصة الذين بينهم درجة من المودة والعاطفة كالزوجين مثلا. وتعتبر مؤشرا لارتياح أو عدم ارتياح المستجيب لمثل هذا التقارب العاطفي.

- **أدوات اسقاطية تركز على الشكل** كما في اختبار الرورشاخ أو المسمى باختبار بقع الحبر Rorschach Inkblot Test الذي يصنف استجابات الأفراد من حيث بعض الأقسام الشكلية، مثل الحركة، والنزعة إلى استخدام الكل أو الأجزاء في بقعة الحبر، وتنظيم الشكل، والحساسية للألوان والظلال. ويندرج تحت هذا النوع من الأدوات عدد من الأساليب مثل:

- **أساليب التداعي** وفيها يطلب من الشخص الاستجابة بأقصى سرعة لمثيرات مثل الكلمات أو الصور أو المدركات ومن أمثلة ذلك اختبارات تداعي الكلمات وإكمال الجمل واختبار الرورشاخ وكشف الكذب.

- **الأساليب البنائيـــة** ويطلب فيها من الفرد كتابة قصة أو رسم صورة دون أي قيد على استجابته ومن أمثلة ذلك اختبار تفهم الموضوع.

أدوات اسقاطية تعتمد على التكميل كما في اختبارات تكملة القـــصص Story Completion Test ومن أكثرها شيوعا قائمـــة روتـــر (Rotter, 1972) للجمل الناقصة التي تشتمل على 40 جملة ناقصة من مثل:

- **أتمنى لو أنني ......،**
- **معظم الناس .....**
- **والدتي ....**

وغالبا ما يكون الزمن المخصص للإستجابة على مثـــل هـــذه الجمل مفتوحا، ويتم تحليل الإجابات واستخدامها بعد ذلك لأغـــراض اكلينيكية وغيرها. وقد تم تقديم هذا الإختبار في ثلاثـــة اختبـــارات، الأول للطلبة من الصف التاسع حتى الثاني عشر، والإختبار الثـــاني للمرحلة الجامعية، والإختبار الثالث للكبار. ويتم تقييم كل إجابة وفقا لمقياس متدرج من 7 نقاط، يتدرج من (1): يحتاج إلى علاج نفسي، إلى (7) متوافق بدرجة عالية جدا.

- **أساليب الإختيار أو الترتيب:** الشخص أن يختار بديلا من عدد من الصور أو البدائل بديلا أكثر توافقا مع محك معـــين مثـــل الاستحسان والفرق ومن أمثلة ذلك اختبار تـــومكنس وهـــورن لترتيب الصور Tomkins-Horn Picture Arrangement Test.
- **أساليب تعبيرية (ألبورت Allport):** وفيها يطلب من الشخص القيام بدور نشط في الرسم والطلاء أو اللعب كوسيلة للتعبيـــر عن شخصيته، وهنا يكون التركيز على النتاج، وأسلوب أو نمط التعبير ولهذه الأساليب قيم علاجية وتشخيصية. إذ أن الشخص

يعبر عن نفسه وكذلك بطريقة تؤثر في تكيفه الشخصي ومن أمثلة ذلك مواقف اللعب مع الأطفال والرسم الشخص وأساليب الطلاء.

- **أدوات اسقاطية تركز على المحتوى والشكل معا:** كما في اختبارات الرسومات مثل اختبار غودانف لرسم الرجل Draw-A-Man-Test عام 1926. وقد تم تطوير الإختبار على يد هاريس عام 1936 وذلك بأن ركز على رسم المرأة إضافة إلى رسم الرجل، وأخذ يعرف عندها باختبار غودانف-هاريس لرسم الشخص: Good Enough-Harris-Draw-a-Person-Test.

- **مقاييس شخصية غير محددة البناء ولكنها واضحة الهدف**

تقع هذه المقاييس في أساليب الملاحظة والمقابلة ومقاييس مفهوم الذات. إذ تعتمد هذه المقاييس في مجملها على دراسة السلوك بملاحظته ميدانيا في مواقف طبيعية، أو إجراء مقابلة وجاهية، كما أن السيرة الذاتية التي يكتبها كل فرد عن نفسه مثال على هذا النوع من المقاييس. وتبرز أهمية هذه الأساليب في أن أساليب قياس الشخصية السابقة تتطلب ورقا وقلما وتستدعي استجابات في مواقف غير طبيعية (أو مصطنعة)، بالرغم من إمكانية ضبطها وتقنينها، إلا أنها تختلف عن مواقف الحياة الفعلية.

أساليب ملاحظة السلوك Observing Techniques: وتعد الخطوة الأولى من خطوات البحث العلمي، وهي وسيلة يستخدمها الإنسان العادي في اكتساب خبراته وجمع بياناته في المواقف موضع البحث مباشرة. ويتميز أسلوب الملاحظة عن غيره من أساليب القياس في كونها تساعد على جمع الاستجابات على طبيعتها (أو على غير طبيعتها) وفي ميادين حدوثها، وفي الأوقات التي ربما تقف المعايير الخلقية حائلا دون

دراستها باستخدام أدوات قياس أخرى كالإختبارات. كما أنها تساعد في دراسة السلوك التلقائي للأفراد.

ويمكن تصنيف الملاحظة حسب طبيعتها في نوعين بسيطة غير مضبوطة أو ملاحظة منظمة وقد تكون فردية أو جماعية، ويمكن أن تكون ملاحظة بالمشاركة أو ملاحظة بغير المشاركة.

أساليب المقابلة Interview Techniques: وهي أداة هامة للحصول على البيانات من خلال المصادر البشرية، وتتكون بأبسط صورها من مجموعة من الأسئلة والفقرات التي يقوم المختبر بإعدادها وطرحها على الشخص موضع البحث ومن ثم يتم تسجيل الاستجابات. وقد شاع استخدام المقابلة في المجالات الاكلينيكية بالرغم من انتشار استخدامها في ميادين أخرى يمكن تطبيق التبادل اللفظي بين شخصين أو طرفين.

وتستخدم المقابلة في فرز الأشخاص لمهام أو مهارات معينة في التشخيص المبدئي حسب درجات المرض النفسي وحالاته ، كما قد تلعب المقابلة دورا استطلاعيا أو تشخيصيا أو علاجيا أو استشاريا أيضا. وتكون المقابلة حرة، أو مقننة، أو بؤرية، أو غير موجهة.

- **مقاييس شخصية محددة البناء لكنها غير واضحة الهدف**

تصنف هذه المقاييس بأنها مقاييس أداء، إذ أنها تهدف إلى الحصول على معلومات تتعلق بسلوك معين في ظروف مقننة، ويمكن تقسيم هذه المقاييس إلى:

- **مقاييس الأنشطة الفسيولوجية** التي تستخدم في تقييم التغيرات التي تحدث في الاستجابات نتيجة للإثارة أو لمواقف ضاغطة. فأنماط الموجات الكهربية للمخ، ونبض القلب، وضغط الدم، ونشاط الغدة الدرقية، وكيمياء الدم ترتبط بالإثارة الإنفعالية واليقظة وتركيز الانتباه، ويختلف الأفراد في النمط المميز لهذه الأنشطة الفيزيولوجية. فجهاز

كشف الكذب Polygraph الذي يستخدم في تسجيل بعض من الأنشطة الفسيولوجية المذكورة أثناء قراءة، أو مقابلة، أو تحقيق معين. أما جهاز رسم المخ الكهربائي (EEG) Electroencephalogram فيستخدم في تخطيط موجات المخ الكهربائية وفي عمليات التقييم النفسي الفسيولوجي في المجالات الاكلينيكية وطب الأعصاب (علام، 2000).

- **مقاييس الإدراك** وتعد حلقة وصل بين مقاييس القـدرات ومقـاييس الشخصية. تم التوصل إليها من خلال تجارب علم النفس المعرفي في مجال المعرفة Cognition والإدراك Perception. فقـد لـوحظ وجود فروق فردية في عملية تجهيـز المعلومـات (Processing) والانتباه Attention. وقد عرفت هذه الفروق بالأساليب المعرفية، وهي فروق في استراتيجيات الضبط، وأطلق عليها روتر (Rotter, 1972) الضبط الداخلي والضبط الخارجي.

ثانيا: تصنيف أدوات القياس حسب الصفة والهدف

تعددت أساليب تصنيف أدوات القياس عموما، وذلك تبعا لاختلاف طريقة الاستخدام أو طبيعة تفسير النتائج أو عدد المفحوصين أو مقدار الزمن المخصص لتطبيق أداة القياس أو شكل فقرات الإختبارات وطريقة تصحيحها. وفيما يلي عرضلبعض طرق تصنيف أدوات القياس حسب عدد من الأبعاد.

*. فقد تصنف أدوات القياس حسب **وجود المثير والاستجابة إلى:**

أدوات إسقاطية projective tools عندما لا يكون أي من المنبه أو الاستجابة محددا كاختبار البقع الحبرية (الرورشاخ). واختبار تفهم الموضوع (TAT) وبعض الإختبارات النفسية الأخرى. **وأدوات محددة البناء Well-Structured** كما في اختبارات التحصيل والقدرات والإستعدادات. وتستخدم عبارات معينة من أجل المساعدة في الوقوف على

مستوى الطلبة في مجالات محددة كمستوى الطلبة في التعبير والكتابة أو غيرها كما وتستخدم هذه الأساليب الإسقاطية أحيانا في استقصاء نوعية وحجم المشكلات التي يواجهها الطلبة في دراستهم أو في حياتهم الأسرية وغيرها.

*. **وتصنف أدوات القياس حسب مستوى الأداء إلى أدوات ذات أقصى أداء maximal performance tools** وهي الأدوات التي يطلب من الممتحن أن يبين أفضل ما لديه من قدرات أو معلومات عند الإجابة عليها، كما في اختبارات المهارات والتحصيل والقدرات والاستعداد. إذ تتضمن هذه الأدوات إجابة صحيحة أو نموذجية وغيرها يعد إجابة خاطئة. وهنا من الممكن تشجيع الطلبة على الدراسة والتعمق لبلوغ درجات تحصيل أفضل على هذا النوع من الإختبارات. **وأدوات ذات الأداء العادي typical performance tools** وهي الأدوات التي يطلب من الممتحن أن يبذل أداء عاديا كما في قوائم الميول والاتجاهات والشخصية وغيرها . إذ لا تتضمن هذه الأدوات إجابة صحيحة أو نموذجية. وهي تعكس سلوك المتعلم في الأوضاع العادية. ولا حاجة إلى تشجيع الطلبة على الدراسة ، بل إلى توخي الدقة والموضوعية في الإجابة.

*. وإذا أخذنا بعين الاعتبار **عدد الأفراد الذين ستطبق عليهم أداة القياس**، فإنه يمكن تصنيفها إلى **أدوات فردية Individualized tests** كما في اختبارات الشخصية والإمتحانات الشفوية والفحص الطبي وبعض اختبارات الذكاء (ستانفورد – بينيه) واختبارات ويكسلر لذكاء الأطفال وأطفال ما قبل المدرسة، والراشدين ، إضافة إلى اختبارات القدرات الفنية وبعض اختبارات المهارات. **والأدوات الجمعية Grouped tests** كما في الإختبارات التحصيلية الصفية (اختبارات كل من منتصف ونهاية الفصل في المستوى الجامعي وامتحانات الثانوية العامة) والإختبارات المقننة

واختبارات القدرات العقلية من مثل اختبار كاليفورنيا للنضج العقلي واختبار أوتيس لينون للقدرة العقلية العامة وختبار بيتا III للذكاء غير اللفظي.

*. ويمكن تصنف أدوات القياس حسب **شكل الفقرات أو طريقة الإجابة** **إلى اختبارات ذات الإجابة المنتقاة Selection-type responses** وتقع في عدة أنواع من مثل اختبارات الإختيار من متعدد، والمزاوجة واختبارات الإختيار من بديلين كالصواب والخطأ أو الإجابة بـ (نعم) أو بـ (لا). أو **اختبارات ذات الإجابة المصوغة Supply-type responses** وتقع في عدة أنواع من مثل اختبارات التكميل واختبارات الإجابة القصيرة والإختبارات الإنشائية ذات الإجابة مفتوحة النهاية (المقالية) والإختبارات المتعلقة بحل المسائل.

*. وتصنف الفقرات حسب **مقدار الزمن المخصص للاستجابة.** فهناك **اختبارات القدرة Power tests** ويكون الزمن المخصص للإجابة عليها مفتوحا نسبيا كما في اختبارات التحصيل والقدرات والإستعدادات. وهنا ليست السرعة في الاستجابة مسألة حاسمة، بل هي أمرا ثانويا. إذ أن الموضوع الهام هنا هو قدرة المفحوص على الاستجابة، وليس سرعته في إنجازها. وخير مثال على ذلك الإختبار المتعلق بفحص طبيب جراح متقدم لشغل وظيفة في مستشفى، فهذا اختبار قدرة ، حيث أن المهم هو إتقان إجراء العلمية الجراحية وليس السرعة في إجرائها. وكذلك **اختبارات السرعة Speeded tests** التي يكون الزمن المخصص للإجابة عليها محددا نسبيا كما في اختبارات المتعلقة ببعض المهارات مثل سباقات الجري والسباحة والقفز أو حتى الطباعة. وهنا تعد السرعة في الاستجابة مسألة حاسمة، إذ أن الموضوع الهام هنا ارتباط أداء المفحوص بسرعة الاستجابة أو زمن الاستجابة المحدد. والمعروف أن فقرات اختبارات السرعة سهلة عموما، إذ يستطيع المفحوص الإجابة على كافة فقرات الإختبار لو يعطى الوقت الكافي.

*. وحسب **طريقة التصحيح**، تصنف أدوات القياس إلى: **موضوعية Objective** وهي أدوات القياس التي لا تتأثر درجات الممتحنين عليها بشخصية المصحح أو وجهة نظره أو خلفيته الاجتماعية أو السياسية. إذ يتم إجراء عملية التصحيح باستخدام مفتاح مثقب لهذا الغرض أو تتم العملية آليا أو باستخدام الحاسوب. وهذا إنما ينطبق على تصحيح الإختبارات ذات الإجابة المنتقاة. وهناك أدوات **ذاتية Subjective**، تتأثر درجات الممتحنين عليها بشخصية المصحح أو وجهة نظره أو خلفيته الاجتماعية أو السياسية. إذ يتم إجراء عملية التصحيح من قبل مصححين ربما يعرفون الممتحنين بشكل أو بآخر. وهذا يعتمد على فهم وتقدير هؤلاء الأشخاص لمستوى أداء الممتحن، و يبدو ذلك واضحا أكثر في حالة الإختبارات ذات الإجابة المصوغة وخصوصا تلك الإنشائية ذات النهاية المفتوحة، وكما في مواضيع التعبير الأدبي وغيرها.

*. وتصنف أدوات القياس حسب **كيفية ظهور الأداء** إلى **مقاييس لفظية Verbal** وهي تلك الأدوات التي يتم استخدام اللغة فيها بشكل رئيس في التعليمات والفقرات ذاتها. وهذه تتطلب قدرة الممتحن على القراءة والكتابة، إضافة إلى أن هذه الأدوات تشيع في المجالات المعرفية والوجدانية وتقل نسبيا في مجال المهارات الحركية. **ومقاييس غير لفظية Non-Verbal** أو ما تدعى مقاييس الأداء العملي، وهي تلك الأدوات التي لا يتم استخدام اللغة فيها بشكل رئيس سواء في التعليمات أو في محتوى الفقرات ومدلولها. وهذه لا تتطلب كثيرا قدرة الممتحن على القراءة والكتابة كما في الأدوات التي تستخدم مع الأطفال المعاقين أو الأطفال الصغار أو الأميين، وهي فقرات تشتمل تكميل الصور وترتيبها وتجميع الأشياء وتصميم المكعبات وتتبع المتاهات. ولا يخفى أن هناك بعض فقرات اختبارات الذكاء الفردية التي تتضمن بعض الفقرات المتعلقة بالاستعداد المكاني والحركي وجوانب نفسية وحركية. وجدير بالذكر أن الإختبارات

ذات الطبيعة اللفظية أكثر حاجة إلى التقنين (اشتقاق المعايير المناسبة) من تلك غير اللفظية.

*. تصنف أدوات القياس حسب **الجهة التي تقوم على بناءها وتطويرها إلى اختبارات مقننة أو منشورة Standardized or Published tests** وتدعى أحيانا اختبارات رسمية formal **واختبارات من إعداد الشخص Man-made tests** أو التي تعرف أحيانا **بالأدوات غير الرسمية Informal tests.**

*. تصنف أدوات القياس حسب **درجة تحديد المجال الذي يتم قياسه** إلى **اختبار عينة Sample test** وهو الإختبار الذي يتم تكوين فقراته عشوائيا كعينة ممثلة لمجمل الفقرات التي تشكل مجال السمة موضع الاهتمام، كما في الإختبارات التحصيلية التي يكون فيها ذلك المجال محددا. أي أن هذا الإختبار يقيس عينة من السلوك يستطيع المقوم التعرف على مدى تمثيل هذه العينة من الفقرات المجال السلوكي ككل. **واختبار كمؤشر Sign test** وهو الإختبار الذي يتم تكوين فقراته بطريقة لا يمكننا القول بأنها تمثل المجال موضع الاهتمام، إذ أن المجال في هذه الحالة لا يمكن تحديده تماما ويعرف أحيانا بأنه المجال مفتوح النهاية. ويشمل ذلك مجال قياس القدرات العقلية كالذكاء وبعض السمات النفسية.

وجدير بالذكر أن هذا النوع من الإختبارات التي قد تستخدم في المقارنة بين مجموعات من الأشخاص في سمة معينة تختلف عما يقيسه الإختبار ذاته. فقد يعد اختبار لتذكر المفردات اللغوية (المجال محدد) على أنه اختبار مؤشر عندما يستخدم في التمييز بين المتخلفين عقليا والعاديين.

*. وتصنف أدوات القياس حسب **درجة تشابه السلوك الذي يقيسه الإختبار والسلوك المتنبأ به** إلى **اختبار ممثل Represent** وهو الإختبار الذي يشبه السلوك الذي يقيسه في مرحلة ما ذلك السلوك المنوي التنبؤ به في مرحلة دراسية لاحقة. فمثلا يعد امتحان الرياضيات في المرحلة الثانوية

اختبارا ممثلا إذا استخدم في التنبؤ بمستوى أداء الطلبة في الرياضيات في المرحلة الجامعية. وهناك **الإختبار المتنبىء Predict** ، وهو الإختبار الذي لا يشبه السلوك الذي يقيسه في وقت معين ذلك السلوك المنوي التنبؤ به في مرحلة دراسية لاحقة. فمثلا يعد اختبار الذكاء المستخدم لأغراض القبول في تخصص معين أو شغل مهنة أو وظيفة ما اختبارا متنبئا إذا استخدم في التنبؤ بمستوى أداء هؤلاء المرشحين المقبولين في دراسة التخصص أو شغل المهنة أو الوظيفة موضع البحث.

وتصنف أدوات القياس حسب **طريقة تفسير نتائجها** إلى **اختبارات معيارية المرجع Norm-Referenced-Tests (NRT)** وهنا يتم تفسير مستوى أداء الممتحن عليها بالمقارنة مع متوسط أداء المجموعة التي ينتمي إليها وهذه المجموعة التي تعرف بالمجموعة المعيارية norm group التي قد تكون صفا دراسيا أو مجموعة مهنية معينة أو مدرسة أو منطقة تعليمية بكاملها. وربما يختلف مستوى أداء الممتحن حسب المجموعة التي ينتمي إليها، فقد يكون ترتيبه الخامس من بين طلبة صفه، وقد يكون العشرين على مستوى مدرسته، كما أنه من الممكن أن يكون ترتيبه المائة بالنسبة لمنطقته التعليمية وهكذا. كما يمكن تفسير درجة الممتحن بمقارنتها مع متوسط أداء المجموعة التي ينتمي إليها وفي ضوء مقدار الانحراف المعياري لمجمل الدرجات، فقد يمكن القول بأن العلامة 90 مثلا تزيد عن المتوسط الحسابي للمجموعة (70) بمقدار أربعة انحرافات معيارية (5)، كذلك بالنسبة للدرجة 50 التي تقل بمقدار أربع انحرافات معيارية عن ذات الوسط.

يشيع استخدام هذا النوع من الإختبارات في أغراض قياس مستوى التحصيل الدراسي في مراحل دراسية متقدمة، وفي أغراض التعيين والتقويم الختامي. كما أن الفقرات من نوع الإجابة المنتقاة هي الأكثر استخداما في هذا النوع من الإختبارات. وفي الغالب تستطيع هذه

الإختبارات أن تغطي محتوى دراسي أكبر بعدد قليل نسبيا من الفقــرات. وهي تقيس أهدافا أقل تفصيلا من حالة الإختبارات من نوع محكية المرجع. وهناك **اختبــارات محكيــة المرجــع Criterion-Reference-Test (CRT)** وفيها يتم تفسير مستوى أداء الممتحن عليها بالمقارنة مع مستوى أداء معين يتم تحديده مسبقا من قبل واضع الإختبار أو المؤسسة التي يــتم تطبيق الإختبار لصالحها، دون النظر إلى أداء المجموعة التي ينتمي إليهــا الممتحن. كأن يكون المستوى الذي يحتكم إليه هو أن يحل الممتحن 85 % من مجمل المسائل المعطاة، أو أن يقفز المتسابق مثلا 1.5 متــرا ، أو أن يقوم المتسابق بطباعة 45 كلمة عربية في الدقيقــة أو طباعــة 55 كلمــة إنجليزية في الدقيقة طباعة صحيحة. ويشيع اســتخدام هــذا النــوع مــن الإختبارات في أغراض قياس مستوى التحصيل الدراسي في كافة المراحل الدراسية، مع أن التركيز في شيوعها في المرحلة الابتدائية. إضــافة إلــى شيوعها في أغراض التشخيص وعمليات التقويم البنائي أو المرحلي. كمــا أن الفقرات من نوع الإجابة المصوغة هي الأكثر استخداما في هذا النــوع من الإختبارات.

وفي الغالب تستطيع هذه الإختبارات أن تغطي محتوى دراسي ضــيق نسبيا بعدد كبير من الفقرات. وهي تقيس أهدافا أكثر تفصيلا من حالــة الإختبارات من نوع معيارية المرجع.

وبشكل عام تعتمد **نوعية Quality** الإختبارات معيارية المرجــع علــى مقدار التباين بين درجات الممتحنين، وهذا يــشير إلــى أن ثبــات هــذه الإختبارات يعد مسألة في غاية الأهمية. بينما يكون الاهتمام أكبر بمؤشرات صدق الإختبار في حالة اختبارات المحكية المرجع وقدرتها على التحــسس لنوع التدريس ومدى قدرة فقــرات الإختبــارات علــى قيــاس الأهــداف الموضوعة. ويهتم مصممو الإختبارات عموما بنوع الإختبار المطلوب ليتم مراعاة ذلك في خطوات بناء الإختبارات.

يبدو أنه مازال هناك بعض من الإرباك والتداخل في الأدبيات المتعلقة بتعريف كل من مفهومي NRT و CRT. فقد يعتقد البعض عند استخدامهم للمصطلح NRT أنه يتحدث عن الإختبارات المقننة، والعكس صحيح، رغم إن المصطلحين غير مترادفين). وعلى نفس المنوال، يستخدم بعض الاختصاصيين مصطلح CRT بطرق متباينة.

أما بخصوص درجة معياري المرجع، فإن الفرق هو في المعنى المطلق إزاء المعنى النسبي. فقد يميل الأفراد إلى تفسير الدرجات في الإختبارات كما لو كانت ذات معنى مطلق. فعلى سبيل المثال، غالباً ما يحدد التربويون درجة نجاح على أنها (60 %). وهذه تعمم أيضا على اختبارات عدة، وعليه، فإذا أحرزت (سميرة) (62 %) في اختبار التهجي، و(58%) في اختبار الرياضيات، فإن المرء يستنتج أن سميرة كانت قد أظهرت أداء في اختبار التهجي أفضل منه في اختبار الرياضيات. ومشكلة هذا الاستنتاج أنه يفترض معنى مطلقاً للدرجات. أما إذا استنتجنا أنها نجحت في أحدهما وفشلت في الآخر، فإنه يستوجب النظر في أمرين هما:

(أ) مقدار الأداء اللازم لإحراز درجة النجاح،

(ب) إن هذا المقدار متساو في كلا الحالتين.

والأمران عرضة للتساؤل. فنظراً لكون الكثير من الافتراضات المشكوك بها مرتبطة بالطريقة المطلقة للتفسير آنفة الذكر، قام المختصون في القياس بتطوير فكرة التفسير المعياري المرجع (NRT)، وذلك بإضافة معنى للدرجة من خلال مقارنتها بدرجات الآخرين في مجموعة مرجعية (أو معيارية). فعلى سبيل المثال، إن الدرجة (62 %) التي حصلت عليها (سميرة) في التهجي قد نضعها في مستوى النسبة المئوية الخمسين بالمقارنة مع زملائها، أو بعبارة أخرى، كان أداؤها أفضل من أداء خمسين بالمائة من زملائها. بينما تضعها درجتها في الرياضيات في الموقع أن أداءها أفضل من أداء ثمانين بالمائة من زملائها.

يرى البعض أن الطريقة المعيارية المرجع لا تزودنا بالمعلومات المهمة فعلاً بخصوص ما يتعلمه الطلبة وعن مستوى تعلمهم. فهم يرغبون باستخدام الطريقة المطلقة للتفسير وقد أوجدوا مصطلحاً لها حيث أطلقوا عليها (التفسير المحكي المرجع).

إذا ما قمنا بتفسير درجة أحد الطلبة بمقارنتها مع درجات الآخرين (الذين نطلق عليهم المجموعة المعيارية)، فسوف يكون هذا معياري المرجع. أما إذا فسرنا أداء الطالب من خلال مقارنته بإحدى الحقول السلوكية المتخصصة أو بمحك معياري معين، فسوف يكون هذا محكي المرجع. ولتوضيح الفرق، نستطيع القول إن بؤرة اهتمام الدرجة المعيارية تنصب على السؤال التالي: (كم هو عدد زملاء (سعيد) الذين كان مستوى تحصيلهم أقل من مستوى سعيد ؟). أما أساس اهتمام الدرجة المحكية المرجع فينصب على ما بوسع (سعيد) أن يفعل. وبالتأكيد فإن باستطاعتنا تفسير درجة اختبار واحدة وفقاً للطريقتين.

في المعياري المرجع، قد نستطيع قول العبارة التالية: قام (سعيد) بأداء أفضل من ثمانين بالمائة من الطلبة في اختبار في عملية الجمع للأعداد الصحيحة. أما في المحكي المرجع، فيمكننا القول بأن سعيد حصل على (70 %) من الفقرات الصحيحة في اختبار عملية الجمع للأعداد الصحيحة. كما يمكننا أن نضيف معنى آخر لهذه العبارة من خلال القول ما إذا كانت نسبة (70 %) حسب اعتقادنا، نسبة غير دقيقة، أو على مستوى ضعيف من الدقة، أو ممتازة، أو أي شيء آخر.

هنالك بعض الجدل بخصوص ما إذا كان CRT يحمل معه أي معيار ضمني أو مجموعة معايير ضمنية. وهذا يعني: هل بإمكاننا إرجاع الأداء إلى نقطة فاصلة (أو إلى مجموعة من الدرجات الفاصلة)؟ وهذا يعتمد على مفهومنا عن المحك (أو المعيار).

كان غلاسر (Glaser, 1963) من الأوائل الذين استخدموا مصطلح CRT، حيث اقترح ذات مرة أن: "المقاييس محكية المرجع تدل على محتوى الخزين السلوكي والتوافق بين ما يفعله الفرد والبنية التحتية للتحصيل" . كما ذكر في نفس المقال: "أننا نحتاج إلى أن نخصص المستويات الدنيا للأداء بشكل سلوكي". وبعد ذلك، عرف غلاسر ونيكتو (Glaser & Nitko, 1971) الإختبار المحكي المرجع على أنه: "ذلك الإختبار المصمم خصيصاً لكي يعمل على إخضاع المقاييس القابلة للتفسير لبعض المعايير الأدائية المتجددة".

وعلى أية حال، فإن المنادين بهذا النوع من القياس، كما هو واضح، لم يقصدوا أن المعيار درجة فاصلة، على الرغم من سوء الفهم المتكرر الذي يعاني منه CRT.

والجدير بالقول، فقد استمر نيكتو (Nitko, 1980) في استخدام تعريفه السابق، لكنه أوضح أنه ينبغي عدم خلط معنى المحكي المرجع مع فكرة وجود درجة نجاح أو درجة فاصلة. أما بوبهام الذي يعتبر أحد أبرز مؤيدي المقاييس محكية المرجع، فقد أشار إلى أن المحك هو معيار الأداء (Popham, 1969). وعليه، فقد ذكر " أن المحك هو تفسير لمستوى كفاءة الشخص الممتحن ولا يتمخض عن شيء أكثر من الممارسات الإختبارية التقليدية". ومهما تكن الظروف، فإن الجميع يتفقون على أنه بوجود تفسير محكي المرجع للدرجات، يكون الاهتمام منصباً على (ما يمكن للطالب أن يفعل) وتكون المقارنة مع مجال سلوكي. كما إن مسألة ضرورة وجود معيار ضمني للكفاءة أو درجات فاصلة هي مسألة جدلية. ونعتقد أيضا، أن أغلب المستخدمين يفكرون بالمرجعية المحكية للدرجات بهذا الشكل بسبب الإرتباط الوثيق في أذهان الغالبية بين CRT والإتقان أو Mastery أو اختبارات الكفاءة الدنيا.

وبالرغم من وجود بعض الخلافات حول ناحية الكفاءة، يتفق خبراء القياس عموماً على وجود فروقات أساسية بين التفسير معياري المرجع للدرجة والتفسير محكي المرجع لها. ومع ذلك، هنالك العديد من الخلافات والتناقضات حول الفوارق بين الإختبارات المعيارية المرجع والإختبارات المحكية المرجع. وتوحي التعاريف التي مر ذكرها بالإختبارات المحكية المرجع أنها تصمم بشكل يتيح الخروج من نتائج أسئلة الإختبار بما يخدم الحقل بمجمله.

ومهما يكن من أمر، يرى بعض المؤلفين أن مصطلح مجالي المرجع Criterion-Domain أكثر دقة، حيث لا يحمل هذا المصطلح أي إشارة إلى درجة فاصلة أو معيار، وهذا يناسب هؤلاء الذين يرغبون بحذف هذا المعنى من المحكي المرجع. إلا أن البعض الآخر استمر باستعمال مصطلح (المحك). ونقطة خلافهم بهذا الشأن هي أنه:

"حتى وإن كان هناك الكثير من الإرباك والقليل من الوضوح في أذهان العديد من التربويين بخصوص فكرة القياس، فإن من المسلم به عموماً أن التربويين قد وجدوا في القياس محكي المرجع طريقة جديدة للتقييم قدمت فوائد في أغراض معينة أكثر من تلك التي تقدمها إستراتيجيات القياس التقليدية."

ويبدو، أننا سنبقي على مصطلح يتقبله التربويون رغم ما يسببه لهم من إرباك. ويعزى الإرباك الحالي الخاص بالمصطلحات والتعاريف نوعاً ما إلى سوء فهم صدق المحتوى. إذ يرتبط صدق المحتوى أساساً بمدى دقة فقرات الإختبار في تمثيل المجال الذي يراد عمل الاستنتاجات لصالحه.

يقول بعض مؤيدي الإختبارات المحكية المرجع ــ أو هكذا بدا لهم ــ بأن القياس المعياري المرجع يفتقر إلى مقارنة الطلبة وليست له القدرة على توفير أي معلومات بخصوص ما يستطيع الطالب أن يفعل كما لو كانت مثل هذه المقارنة غير مبنية على أي محتوى.

وغالباً ما يكون الجدل حول المحاسن النسبية لاختبارات التحصيل المقننة التقليدية والتي: "يشير إليها الآخرون على إنها اختبارات معيارية المرجع." واختبارات مجالية المرجع مناسبة لأهدافها بفاعلية. ويعتقد البعض أن الدرجات الناجمة عن الإختبارات التقليدية لا يمكن أن تكون مجالية المرجع لعدم تعريف المجال بشكل واضح، ولأن الفقرات ليست عينة عشوائية من المجال. ورغم أن الناشرين لبعض الفقرات المقننة يصفون المحتوى بشكل عام وواسع جداً، يقوم بعضهم بتوفير مخططات تمهيدية ومفصلة. ومن أجل أن يكون الإختبار ذا معنى، يجب ربط كل نوع من درجات الإختبار بمحتوى الإختبار، وبدرجات الممتحنين الآخرين (Ebel, 1962).

استخدامات القياس معياري المرجع

ارتكزت معظم الإختبارات الفعلية وكذلك نظرية الإختبارات التقليدية على الأدوات معيارية المرجع. وهذه الطريقة مفيدة في اختبار الاستعداد، حيث نرغب في الحصول على تكهنات تفاضلية. كما أنها مفيدة جداً في اختبارات التحصيل. إذ أن دور أداة القياس مهم جداً في إعطائنا تقسيماً طبقياً على ضوء التحصيل الذي نقيسه والذي يجب أن يكون ثابتاً وصادقاً قدر المستطاع ولأغراض متعددة. وسوف يتعلم الطلبة مقاديراً متباينة من المواضيع حتى بطريقة تعلم الإتقان. وقد يحدث أن يتعلم كل الطلبة، أو على الأقل نسبة عالية منهم، جزءاً هاماً ومؤثراً من أهداف المعلم والتي يمكن تصنيفها تحت إتقان أساسيات المقرر الدراسي. لكن بعض هؤلاء الطلبة يتعلمون عادة أكثر من غيرهم، ويبدو من الضروري توظيف تقنيات ووسائل قياس لتحديد هؤلاء الطلبة. فالطلبة، بالدرجة الأولى يتوقون للتميز عن الآخرين بإنجازاتهم التي تفوق الحد الأدنى، وهذا من حقهم. فإذا ما أعطينا اختبار إتقان بمستوى مقبول فقط (الحد الأدنى فقط)، فإن هؤلاء الطلبة الذين يحصلون على مستوى أعلى سيفقدون حافزاً داخلياً مهماً للتعلم.

وربما كانت أهمية الإختبارات المعيارية في عملية اتخاذ القرار تشكل سبباً أكثر أهمية من مجرد التركيز أو معرفة الطلبة لمستوياتهم. فكثيراً ما يرغب الطلبة، من أجل التخطيط الوظيفي أو التربوي، في معرفة مستوياتهم مقارنة بالآخرين ذي نفس الخطط. ويشكل المعياري المرجع أهمية كبيرة في انتقاء القرارات أيضاً. إذا ما أتقن اثنان من الأطباء حقل الجراحة، إلا أن أحدهما أتقنه بشكل أفضل، أيهما تختار ليقوم لك بعملية جراحية؟ إذا ما أتقن معلمان أساسيات التعليم، لكن أحدهما كان معلماً أفضل، أيهما نختار لمدارسنا؟ إذا ما أتقن طالبان المقرر المطلوب في مادة الجبر، إلا أن أحدهما تعلمه بشكل أعمق، أيهما يجب أن يلقى تشجيعاً أكثر ليستمر في دراسة الرياضيات؟ من المحتمل أننا جميعاً نتفق على الإجابات. مع ذلك، إن نحن لم نوظف تقنيات القياس التي تتيح لنا التمييز بين الأفراد، فلن يكون بمقدورنا أن نتخذ قرارات كالتي ذكرت آنفا. ومن المؤكد أن المقاييس معيارية المرجع هي الأكثر فائدة في قرارات انتقاء الحصة المحددة. على سبيل المثال، إذا كان لدينا عدد محدد من القبولات في مدرسة إعداد الطيارين، ستقوم المدرسة باختيار أفضل المتقدمين حتى وإن كان الجميع أعلى من بعض مستويات الإتقان.

ولأن الإختبارات المعيارية المرجع المقننة تكون عادة أوسع في محيطها من الإختبارات المحكية المرجع، فإنها أكثر فائدة في إعطاء النظرة الشمولية العامة لمستويات التحصيل في مادة معينة، كما أنها أفضل في تأشير التقدم العام الذي يحرزه الطالب، أو الصف، أو المدرسة. وبالرغم من أن بعض الأفراد يعتقدون أن الإختبارات المعيارية المرجع العامة لا تتأثر بالتعليم في الواقع، نرى أنها تبدي تطوراً ملحوظاً من مستوى دراسي إلى مستوى دراسي آخر. وفي أي اختبار، مهما كان واسعاً أو ضيقاً في محيطه، تكون الدرجة إما معيارية المرجع أو محكية المرجع.

تعتبر الإختبارات المعيارية المرجع عنصراً هاماً في برنامج التقـويم التربوي. لقد ذكرنا أن الإختبارات المحكية المرجع تكون عادة أضيق فـي محيطها من الإختبارات المعيارية المرجع. ويرى الـبعض هـذا المحـيط الضيق على أنه مفيد جداً في برنامج التقويم، ويمكننا بناء اختبار محكـي المرجع كهدف لبرنامج معين لرؤية ما إذا كانت تلك الأهداف قد تحققت أم لا. وفي تقويم برنامج معين، علينا أيضاً معرفة مدى فاعلية هذا البرنـامج إذا ما قورن مع برامج أخرى. وبدون تحديـد عـشوائي للطلبـة للقيـام بالبرنامج (ونادراً ما يكون هذا ممكناً في المدارس)، تحتاج المقارنـة لأن تتم من خلال بعض المناهج المعيارية المرجع والتي تقارن أداء الطلبة في البرنامج مع أداء مجموعة معيارية. إضافة إلى ذلك، قد لا يكـون ضـيق الإختبار المحكي المرجع فائدة مطلقة على الدوام، ففـي بعـض الحـالات نرغب في تقويم نتائج واسعة.

استخدامات الإختبارات محكية المرجع

ينبع التأييد للقياس المحكي المرجع بالدرجة الأساس مـن التأكيـد علـى الأهداف السلوكية، ومن تسلسل وتفريد التعلـيم، ومـن تطـوير المـواد المبرمجة، ومن نظرية التعلم القائلة بأن بوسع أي شـخص أن يـتعلم أي شيء إذا ما منح الوقت الكافي من الاهتمام المتزايد، وأيضاً من الاعتقـاد بأن المعياري المرجع يؤدي إلى تشجيع روح المنافسة غير السليمة وغيـر الصحيحة، وبأنه (أي المعياري المرجع) يـسيء لهـؤلاء الطلبـة الـذين يحرزون درجات واطئة، حيث إنه يؤثر سلباً على مفهومهم وفكرتهم عـن ذواتهم. كانت أهم استخدامات القياس محكـي المرجـع مقتـصرة علـى اختبارات الإتقان داخل الصف، وعلى اختبارات المهارة الدنيا كالتخرج من المدرسة الثانوية، على سبيل المثال، وعلى اختبـارات التـرخيص (مـنح الإجازات). يعد اختبار الإتقان نوعاً خاصـاً مـن الإختبـارات المحكيـة المرجع. فالإتقان، كما تدل المفردة، مرتبط بموقف مثل: إما أن الـشخص

الفلاني قد حقق أو أتقن الهدف (أو الأهداف) بشكل مقنع، أو لم يحقق. كما يستطيع الإختبار محكي المرجع أن يقيس درجات الأداء.

وتستخدم اختبارات الإتقان في برامج التعليم الإفرادي أو في نموذج لإتقان التعلم (Bloom, 1968). إذ تتألف هذه البرامج من وحدات، أو وحدات قياسية ينظر إليها على الأغلب بشكل هرمي، ويعتمد كل منها على هدف سلوكي واحد أو أكثر، ويطلب من كل فرد أن يعمل في الوحدة (التعليمية) حتى يحقق مستوى تخصص أدنى من التحصيل. ثم يعتبر ذلك الفرد متقناً ومسيطراً على تلك الوحدة التعليمية. وفي مثل هذه البرامج لا تعتمد القرارات التعليمية حول الطالب على مقارنة أدائه مع أداء غيره. أما إذا أدى بشكل دقيق تجاه الأهداف، يقتضي القرار انتقاله إلى الوحدة الدراسية التالية. أما إذا لم يفعل، فعليه إذن أن يعيد دراسة المادة التي غطاها الإمتحان (ربما باستخدام طريقة مختلفة) حتى يؤدي أداءً دقيقاً، وهذا يعني إتقانه للمادة.

وإذا كانت الإجراءات التعليمية منظمة بشكل يجعل الوقت هو البعد الذي يتغير (المتغير) وكانت درجة السيطرة أو الإتقان ثابتة، فهذا يجعل استخدام اختبارات الإتقان أكثر شيوعاً مما هي عليه الآن. ويتطلب اختبار الإتقان من القائم على الإختبار أن يضع درجة فاصلة. أي يجب أن يكون هناك أساس منطقي دقيق ومنهج لاختيار تلك النقطة وليس هناك من معلومات مفيدة جداً يمكن الحصول عليها تخص درجة الكفاءة فوق أو تحت الدرجة الفاصلة.

من الإستخدامات المتعلقة بالإختبارات المحكية المرجع اختبار الكفاءة الدنيا. إذ أن اختبار الكفاءة الدنيا هو أحد الحقول التي توضع فيها الدرجات الفاصلة. أما الدرجات التي تزيد على الدرجة الفاصلة فلا يتم اعتمادها عند اتخاذ قرارات التطوير والتخرج والترخيص.

تعتبر اختبارات الترخيص Licensing للمعلمين (اختبارات التخــرج وإجازة ممارسة المهنة) اختبارات محكية المرجع. والغايــة مــن اختبــار الترخيص هي التأكيد على حماية الصحة العامة والأمــان والرخــاء فــي المجتمعات. لذا فإن هذه الإختبارات تساعد في تحديد مــا إذا كــان لــدى الأفراد كفاءة دنيا لممارسة مهنتهم، وعليه يجب تأســيس درجــة قطعيــة فاصلة.

كما أن استخدامات التعليم الموجه فردياً أو نموذج الإتقان في الــتعلم واختبار الكفاءة الدنيا والترخيص ليست بالاستخدامات الوحيــدة للمقــاييس المحكية المرجع، حيث يستطيع الفرد استخدام مثل هذه المقاييس في تقــويم البرامج التعليمية واتخاذ القرارات الخاصة بها. فمن أجل تحديد ما إذا كانت بعض المعالجات أو الطرائق التعليمية ناجحة أم لا، لا بد أن تكــون لــدينا معلومات حول النتائج التي تخص بعض الأهداف المحــددة التــي صــمم البرنامج من أجل تدريسها. فقد لا يوفر القياس الذي يقــارن الطلبــة مــع بعضهم البعض (معياري المرجع) معلومات هامة كتلك التي يوفرها مقياس يقارن بين أداء كل طالب والأهداف المرسومة (محكي المرجع).

كما توفر القياسات المحكية المرجع، أيضاً، بعض الفوائد لعملية صنع القرار التعليمي داخل الصف، حيث إن تشخيص بعض الصعوبات المحددة المصحوب بتوصية لنوع من المعالجة التعليمية أمر ضروري في التعلــيم، سواء استخدم المعلم طريقة الإتقان في التعليم أم لم يستخدمها.

ولأن الإختبارات محكية المرجع عادة ما تكون أضــيق فــي ســعة منظورها، فقد يكون هناك بعض الفقرات الكافية ذات هدف معين من أجل استخلاص بعض الاستنتاجات التي تخص الأداء العام للفــرد فــي ذلــك الهدف. ولا يمكن القيام بذلك بشكل نموذجي في الإختبار المعياري المرجع بسبب قلة عدد الفقرات في كل هدف مخصص.

وأخيراً، من الممكن أن تكون الإختبارات محكية المرجع مفيدة جـداً في الاستطلاعات الواسعة للإنجازات التربوية، مثل التقييم العـام للتطـور التربوي أو برامج التقييم المحلية أو الإقليمية.

في الختام، يتفق جميع خبراء القياس على أن كل من NRT و CRT ضروريان في بعض الأحيان في صنع القرار الفاعل. أمـا مـسألة متـى تستخدم التفسيرات المحكية المرجع أو المعيارية المرجع، فهذا يعتمد علـى نوع القرارات المزمع اتخاذها. ففي قرارات التعليم والتوجيه، يستخدم كـلا النوعين (معياري المرجع ومحكي المرجع). بينما بالنسبة لقرارات الانتقاء، فيفضل استخدام الإختبار معياري المرجع. وعند التصنيف، يستطيع الفـرد استخدام كلا النوعين. وفي القرارات المتعلقة بمنح الشهادة، يفضل استخدام الإختبار محكي المرجع.

أما القرارات التعليمية فتعتمد على الأغلب على المنـاهج التعليميـة المستخدمة. فإذا ما كان التعليم مبنياً بحيث يكون عامل الوقت هو المتغير، ويبقى الطالب في عمل مستمر حتى يتقن ذلك التعلـيم، علينـا حينئـذ أن نستخدم اختبار الإتقان. ويستخدم هذا النوع من التعلم في أحيان كثيرة فـي التعليم المنفرد. وإذا كان التعليم مبنياً بشكل يكون فيه وقت التدريس ثابتـاً، فإن الطلبة سيحرزون درجات بمستويات متباينة. ويجب علينا أن نحـاول تتبع هذا الإنجاز المتباين عن طريق الإختبار لغرض التمييز، ومع ذلك، قد نحتاج إلى ربط معنى معياري المرجع ومعنى محكي المرجع بالدرجة. أما مسألة اختيار الطريقة التعليمية التي يجب استخدامها فذلك يعتمد على بنيـة وأهمية المادة الدراسية.

هناك بعض المواد الدراسية التي تكون ذات طبيعة تدريجية هرميـة في تركيبها، والتي تجعل من غير المجدي تدريس مفاهيم عالية المستوى ما لم يتقن الطالب بعض المفاهيم الأساسية أولاً. فعلى سبيل المثال، لا يستطيع الطلبة القيام بالقسمة الطويلة ما لم يتمكنوا من الطرح والضرب بمـستوى

أساسي معين (رغم أننا لا نعرف بأي مستوى). إلا أن هذا الحال لا ينطبق على كل المواد الدراسية. حيث أننا في الواقع لا نستطيع التمكن من قراءة كتاب الذكاء العاطفي أو حتى مجرد قراءته قبل القيام بقراءة مبادىء في علم النفس.

وبنفس السياق، قد تكون هناك بعض المهارات التي تتمتع بأهمية بالغة وتجعل من الضروري إتقانها من قبل الطلبة بغض النظر عن الوقت الذي تستغرقه هذه العملية. وتقع ضمن هذا التصنيف معرفة تهجي الاسم مثلاً. ولكن هذا أيضاً لا ينطبق على جميع المواد الدراسية، وكما ذكر إيبل (Ebel, 1969, P:12) "قد نرغب في السماح لأحد الطلبة بأن يستغرق أسبوعاً لكي يتعلم ما يتعلمه طالب آخر في يوم واحد. لكن، قم بجمع هذه الفوارق مع آلاف الأشياء التي يجب تعلمها. هل هناك شخص، أو طالب، أو مدرس، أو جامعة ممن يريد أن يرى شخصاً أفنى ستة عشر أو أربع وعشرين عاماً في دراسة نفس التعليم الابتدائي الأولي الذي يستطيع آخرون إنجازه في ثمان سنوات؟ أيمكن أن يكون هؤلاء الذين لديهم أدنى درجة ممكنة للتعلم ممن يقضون فترة طويلة من حياتهم في محاولة التعلم؟ إن ممارستنا الحالية عكس ذلك تماماً. إن هؤلاء البارعين في التعلم يتخذونه مهنة لهم. أما الذين ليسوا كذلك فيجدون سبلاً أخرى في المهنة والإنجاز والنجاح.

قام غرونلند (Gronlund, 1985, P: 27) بالتمييز بين الأهداف التي يجب إتقانها من قبل جميع الطلبة، وتلك (الأهداف) التي تعطي الدرجة القصوى من التطور. فبالنسبة للأولى، ينصح باستخدام الإختبار محكي المرجع. بينما يستخدم للثانية الإختبار معياري المرجع. وعليه، فإن هناك مجال لاختبار الإتقان (محكي المرجع) ومجال لاختبار التمييز (معياري المرجع) في سبيل صنع القرارات التعليمية. وقد تكون اختبارات الإتقان أكثر أهمية في المراحل الأولى المبكرة من المراحل المتقدمة في المدارس.

وأخيراً، لا بد لنا من أن نذكر مرة أخرى بأن العديد من الإختبارات تكون معرضة أو بالأحرى طيعة للتفسيرات المعيارية المرجع والمحكية المرجع على السواء. فقد ذكر ناشرو بعض الإختبارات المقننة، على سبيل المثال، درجة معيارية المرجع في كل اختبار فرعي ودونوا داخل كل اختبار فرعي ما إذا كان الطالب قد أجاب على كل فقرة بالشكل المطلوب، إضافة إلى نسبة الطلبة في الصف، والمنطقة التعليمية، أو في المجموعة المعيارية الوطنية الذين توصلوا إلى الإجابة الصحيحة على الفقرة. كما أن إحصائيات الفقرة هذه يتم اختصارها دائماً ببعض الفقرات الخاصة بكل هدف. وبشكل مختصر، يبين الجدول التالي مقارنة بين نوعي الإختبار في ضوء بعض النقاط:

مقارنة بين الإختبار معياري المرجع والإختبار محكي المرجع

وجه المقارنة	معياري المرجع NRT	محكي المرجع CRT
تخطيط الإختبار	يتم تخطيط الإختبار اعتمادا على وصف عام للمهارات والموضوعات التعلمية التي يتضمنها المحتوى الدراسي.	يتم تخطيط الإختبار في ضوء أهداف محددة و ومفصلة ومصاغة سلوكيا، بحيث يتم اشتقاق فقرة أو أكثر لكل هدف.
تحضير فقرات الإختبار	تبنى الفقرات بهدف تعظيم التباين بين الممتحنين.	تبنى الفقرات لقياس كفاية المتعلم في مهارة محددة.
فقرات الإختبار	يمكن استخدام جميع أنواع الفقرات.	يمكن استخدام جميع أنواع الفقرات.
مستوى صعوبة الفقرات	عموما تكون الفقرات على مستوى متوسط من الصعوبة (0.30-0.70) أي أن 30 – 70 % من الممتحنين يستطيعون الإجابة عليها صحيحا.	تكون الفقرات على مدى واسع من الصعوبة. ولكنها تكون الفقرات من نوع التمكن، أي أن معظم الممتحنين يستطيع إجابتها بعد التدريب المناسب والفعال.

محك التمكن	لا يوجد محكات تمكن.	تتضمن الأهداف السلوكية محكات الأداء المقبول.
استخدامات الإختبار	1. يساعد في تحديد مستوى الممتحن جماعة معيارية. 2. يستخدم في انتقاء الأفراد في برامج معينة. 3. يساعد في تحديد مستوى التحصيل في مجال تعلمي محدد.	1. يحدد مستوى أداء الممتحن بالنسبة لمستوى أداء محدد (للتعرف على ما يعرف أو ما يمكنه فعله). 2. يساعد في تشخيص صعوبات ومواطن ضعف المتعلمين أو المتدربين. 3. يساعد في تقييم البرنامج التدريسي أو التدريبي.
الإختبارات المستخدمة	يتوفر منه عدد كبير يقيس الأداء في مناهج دراسية ومجالات عقلية و تربوية ونفسية مقننة.	يزداد عددها تدريجيا، ومعظمها يركز على قياس مهارات أساسية. وتصنف إلى: (1) اختبارات مرجعية الهدف، (2) مرجعية المجال، (3) اختبارات تمكن.
تفسير نتائج الإختبار	بعد حساب درجة الممتحن ، يتم تفسيرها في ضوء موقعها بالنسبة لدرجات المجموعة التي ينتمي إليها. فنقول إن درجة المفحوص تفوق أو تنخفض عن متوسط درجات المجموعة بمقدار عدد من الانحرافات المعيارية.	يتم تقييم نجاح أو عدم نجاح الممتحن على أساس وصوله إلى مستوى تمكن معين أو مدى تحقيقه لهدف محدد. فنقول تمكن المفحوص من أن يحل 80 % من مسائل القسمة الطويلة (بواقع قسمة عدد صحيح يتألف من ثلاث منازل على عدد صحيح من ثلاث منازل على الأكثر حلا صحيحا.

مثال يوضح نوعي التفسير وتم اقتباسه بتصرف من نيتكو (Nitko, 2005):

يحمل هذا المثال العنوان "**الفاصولياء في الجرة Beans in Jar**"، وكان يستخدم كعلامة حدية Benchmark للتقييم في مجلس التربية والتعليم في منطقة تورنتو في كندا.

تهدف المسألة إلى معرفة عدد حبات الفاصولياء التي تملأ جرة كبيرة باستخدام فكرة النسبة والتناسب. حيث يتم توجيه سؤال لكل طالب يطلب تحديد عدد حبات الفاصولياء في الجرة بعد أن يقوم الطالب بشرح طريقة الحساب بتفصيل واضح ودقيق، ويتم عرض طريقة الحل وتفسيرها والدفاع عنها أمام المقيم. ويعطى كل طالب عدة أدوات مثل الميزان والمسطرة وإناء شفاف مدرج.

كما يتم إعلام الطلبة بأنه يجوز القيام بوزن كمية محددة واحدة من حبيبات الفاصولياء، إضافة إلى السماح لهم القيام بعد حبات كمية معينة من الفاصولياء. وفي النهاية، يطلب من كل منهم كتابة وعرض الخطوات التي اتبعها في الحل. وفيما يلي مقارنة بين التفسير معياري المرجع الذي يعتمد على تفسير أو وصف موقع الفرد بالنسبة لأداء أفراد مجموعته والتفسير محكي المرجع الذي يعتمد على وصف مستوى أداء الفرد بالنسبة لمستوى محدد ويتألف من عدد من المهارات والعمليات المطلوب منه القيام بها كما في الجدول التالي:

مقارنة بين التفسير معياري المرجع ومحكي المرجع لدرجة على مهمة معينة

التفسير معياري المرجع NRT	الدرجة على المهمة	التفسير محكي المرجع CRT
20 % من الطلبة حصلوا على الدرجة (5) أو أن 80 % منهم كانت درجاتهم أقل من (5).	5	يفهم الطالب المسألة ويبدأ حالا في البحث عن استراتيجية الحل. يراقب خطواته في الحل ويراقب نتائجه. يستخدم المواد والأدوات المتوفرة لديه بدقة ويرصد نتائجه بحذر. جميع العمليات الحسابية دقيقة وصحيحة ومرتبة ترتيبا منطقيا. يبرر الطالب خطوات عمله ويعرض نتائجه بثقة عالية وبأسلوب واضح ومشوق.
19 % من الطلبة حصلوا الدرجة (4) أو أن 61 % منهم كانت درجاتهم أقل من (4) .	4	ربما يبدأ الطالب بدايات خاطئة، ويخضع لمساعدة المدرس أو المقيم في توجيه تركيزه. يستعمل بعض الأدوات بدون داعي. يخلط أحيانا بين مفهومي الحجم والكتلة. يستخدم فكرة التناسب والنسبة بشكل صحيح. مع أنه يبدو مرتبكا أحيانا أو مضطربا. إلا أنه يقوم بالنهاية بعرض نتائجه الصحيحة ويظهر شوقا لعمله وثقة وإعجابا.
20 % من الطلبة حصلوا الدرجة (3) أو أن 41 % منهم كانت درجاتهم أقل من (3) .	3	هناك بعض الالتباس في واحد أو اكثر من جوانب المسألة. لا يضبط التعبير عن وحدات القياس ويقترف بض الأخطاء في العمليات الحسابية. لديه بعض المعرفة في قضايا التناسب ولكنه غير قادر على توظيفها بشكل صحيح. يتعثر في بعض العمليات ومراحل الحل ولكنه بالنهاية يبرر خطواته بثقة جيدة ويعطي إجابات نهائية صحيحة.
24 % من	2	يستطيع الطالب تحديد بداية صحيحة للحل ولكن لا

الطلبة حصلوا الدرجة (2) أو كانت درجات 17 % منهم أقل من (2) .		يمكنه الاستمرار وإكمال الحل. يتعامل مع المواد والأدوات بشكل صحيح ولكنه غير قادر على التوصل إلى استراتيجية للحل. يظهر خلطا في التعبير عن وحدات القياس وتفسير معانيها. يحاول أن يخمن الإجابة. ولكنه يفتقر إلى الثقة في عمله، ويتوق إلى مساعدة كبيرة من المقيم للقيام بالتوصل إلى الحل.
17 % من الطلبة حصلوا الدرجة (1).	1	ربما يعطي الطالب عدد حبات الفاصولياء ، ولكنه لا يبين أية تفاصيل أو خطوات عن طريقة الحل أو حتى التفكير بأي من تلك الخطوات للتوصل إلى العدد المطلوب بشكل دقيق.

استخدام أدوات القياس والتقويم

ليس من الحكمة القول بأن هذا اختبار جيد وذاك مقياس غير ذلك. بل الأصل أن نقول بأن هناك اختبار جيد لتحقيق هدف معين ولكنه غير جيد في قياس أهداف أخرى. لذلك فمن الأهمية بمكان أن يتم التعرف على القرار الذي يتوقع من التقييم تحقيقه (Nitko, 2002). وقد تم تصنيف القرارات التي يمكن أن تستخدم أدوات القياس المساعدة في اتخاذها بفاعلية كما يلي:

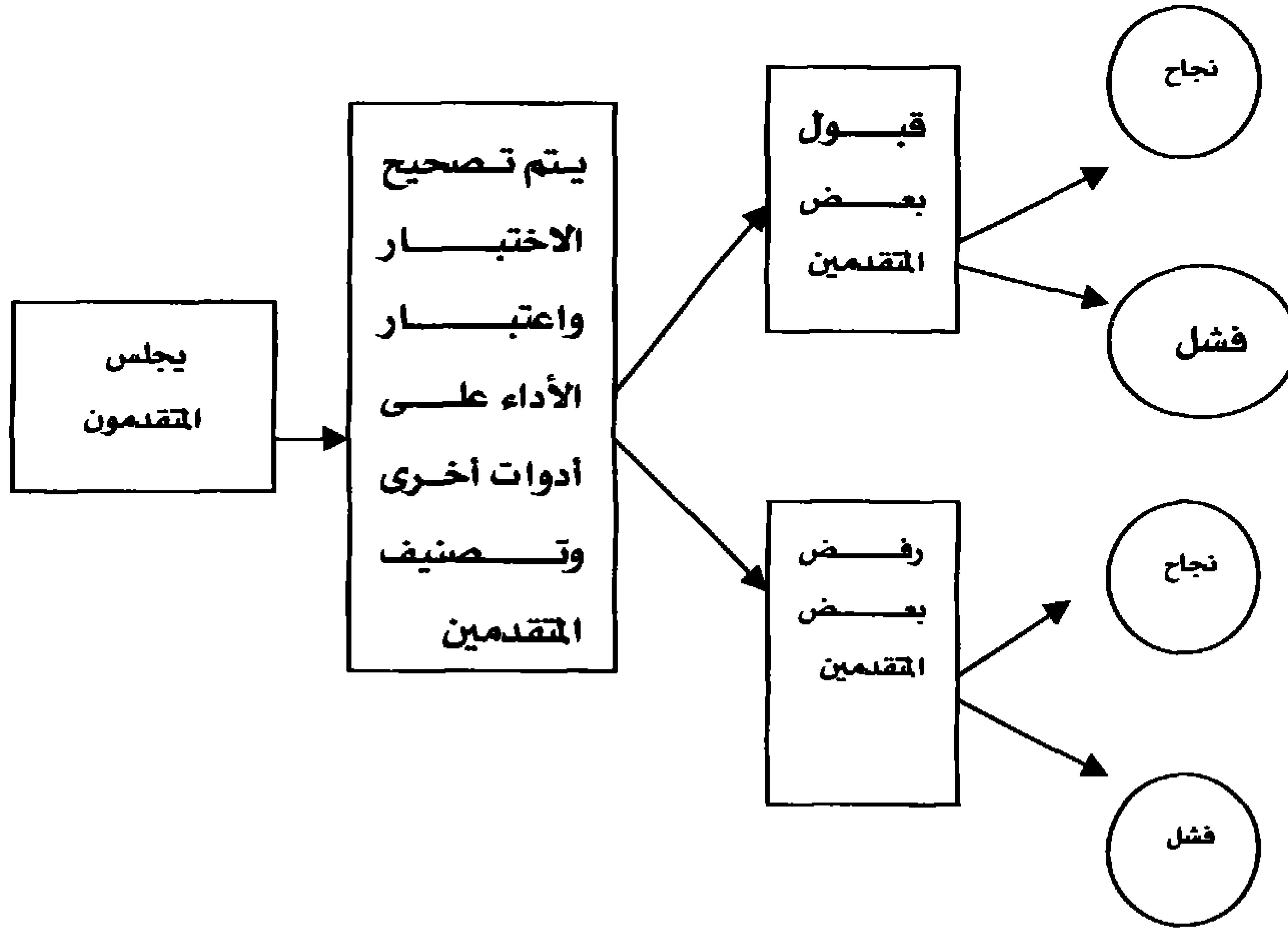
- **قرارات الانتقاء (الإختيار)**
- **قرارات التصنيف**
- **قرارات التسكين**
- قرارات التشخيص والعلاج

- **قرارات الإرشاد والتوجيه**
- **قرارات تطوير البرامج**
- **قرارات تتعلق بتحصيل الطلبة**
- **قرارات تتعلق بناء وتطوير النظريات النفسية والتربوية**

قرارات الانتقاء (الإختيار)

تستخدم أدوات القياس كطريقة للحكم على قبول أشخاص دون آخرين لشغل وظيفة ما أو للالتحاق بمهنة ما أو برنامج معين. فقد يتم استخدام اختبار يتضمن استيعاب بعض المصطلحات الطبية ومهارات الطباعة والاتصال الاجتماعي، وفقرات تتحدث عن الدقة في التعبير، وأخرى عن دفء الاستقبال وود التعامل في تعيين بعض المتقدمين لشغل وظيفة فني استقبال في مستشفى، ورفض آخرين. كما يمكن استخدام أدوات قياس أخرى في قرارات قبول الطلبة المتقدمين في تخصصات جامعية معينة.

وفي العادة، تجري دراسات تنبؤية للحكم على مصداقية قرار الإختيار الذي يبنى بناءا على مستوى الأداء في تلك الإختبارات. فقد يتم حساب معامل الإرتباط بين درجات المتقدمين على اختبار الإختيار ودرجاتهم على محك النجاح في المهنة أو الوظيفة أو البرنامج الدراسي الذي تم قبولهم فيه. ولا شك أن عملية الإختيار في الغالب لا تكون تامة بدون أخطاء، بل هناك هامش ما من الخطأ المتوقع. والشكل التالي يبين مخططا لعملية استخدام اختبار لغرض اختيار متقدمين لوظيفة ما. ويتضح أن قرار القبول أو الرفض فيه نسبة من عدم التوافق.



والجدير بالذكر أن الإختبار الذي يستخدم في اتخاذ قرارات الانتقاء يجب أن يخضع لمراجعة دقيقة دورية، نظرا لخطورة القرار وحساسيته. ويجب تبعا لذلك أن يتم الترويج باتجاه أن قرار الإختيار مرتبط بقرار التنبؤ بالنجاح المستقبلي. لذلك فالأفراد الذين لا يتم اختيارهم ربما يتشككون أو قد يلجأ بعضهم إلى القضاء لشرح شكواهم وعدم رضاهم.

ومما يجعل قرار الإختيار غير فاعل كأن يتم اختيار شخص ليشغل موقعا قياديا في مؤسسة خدمية مثلا، أنه ربما نبني أداة لهذا الغرض ولا نأخذ بعين الاعتبار كافة الخصائص التي يمكن أن يتحلى بها ذلك الشخص ابتداء من العمل اليومي كالالتقاء مع المراجعين من خلفيات اجتماعية متباينة، وانتهاء بتوقيع قرار يقضي بتعيين آخرين أو الاعتذار للبعض، إلى

التوقيع على قرار يقضي بشراء أدوات وآليات تقدر بعدة ملايين من الدولارات.

فقد يتجاهل باني أداة التقييم في التعيينات قدرات الذكاء الاجتماعي أو العاطفي التي تحتل في الواقع حيزا أكبر في العمل اليومي من تلك القدرات التي تحويها اختبارات الذكاء العقلي التقليدية.

قرارات التصنيف والتسكين

يأتي التصنيف بعد أن يتم اختيار أو انتقاء مجموعة من الأفراد للالتحاق بتخصص أو مهنة معينة، كالخدمة العسكرية مثلا (قرار الإختيار). يتم بعد ذلك تصنيفهم في وحدات عسكرية مختلفة (طيران، بحرية، مشاة، مدرعات) ثم يصار إلى تسكينهم في أعمال مناسبة (اتصالات، تدريب، خدمات،.....) داخل هذه الوحدات (Cronbach and Glaser, 1965; Nitko, 1983). ومن ناحية أخرى، قد تستخدم بعض الإختبارات في تصنيف الأشخاص من ذوي الإعاقات مثل صعوبة السمع، والتخلف العقلي، والإعاقة البصرية وغيرها.

ويبدو أن عملية التسكين تتصف بنوع من التفصيل أكثر من عملية الانتقاء سابقة الذكر. فبعد أن يتم اختيار الأفراد المرشحين للقبول في برنامج معين أو في كلية جامعية ما، يتم وضع كل واحد منهم فيما يناسبه من برنامج أو تخصص، أو حتى وضعه في مستوى معين من مستويات البرنامج الذي قبل فيه، بهدف ضمان فرص أكثر للتميز والنجاح. فالتلميذ الذي ربما يتم قبوله (اختياره) في معهد تدريب مهني مثلا، يصار إلى تسكينه (وضعه) في برنامج مهني معين (نجارة، حدادة، كهرباء سيارات،

ميكانيك آليات ثقيلة، ميكانيك أدوات دقيقة). أو ربما يمكن وضعه في مستوى (أول، ثاني، ثالث.....) من برنامج معين (كهرباء سيارات مثلا).

من هنا يمكن أن نفهم أن اختبارات التسكين ربما تلعب دور اختبارات تحديد المستوى. والإختبار الذي يستخدم في اختيار الأفراد ليس بالضرورة أن يستخدم في قرارات التسكين. إذ أن اختبار الانتقاء يركز على مدى القدرة على التنبؤ بالنجاح في البرامج أو الكليات التي تم قبولهم فيها، لكنه لا يفيد كثيرا في التسكين أو التوجيه لقبول المرشحين في برامج أكثر تحديدا وملاءمة.

قرارات التشخيص والعلاج

يلاحظ في بعض الأحيان أن التدريس المقدم لشخص أو لمجموعة من الأشخاص غير ذي جدوى. إذ ربما يكون المتعلم بحاجة إلى مادة تعليمية معينة، أو أن يتم تدريسه بأسلوب ما، وتحت ظروف ومعطيات معينة. إن كشف هذه المعطيات قبيل البدء بعملية التدريس يتم باستخدام بعض الإختبارات والمقاييس التي تعرف بأدوات التشخيص. وربما تعمل أدوات التشخيص إلى الإجابة عن سؤال مفاده: ما الأنشطة التعلمية التي تجعل أداء المتعلم بأفضل صوره (Glaser and Nitko, 1971; Nitko and Hsu, 1974). أما قرار العلاج، فيكون فعالا إذا ما تمت عملية التشخيص بفاعلية كافية.

والجدير بالذكر أن هناك تداخل ملحوظ بين قراري التسكين والتشخيص خاصة عند الحديث عن الإطار التربوي. إذ أن كلا من القرارين يسعى إلى طرح عدد كبير نسبيا ومتنوعا من الفقرات تغطي كافة ميادين الأداء المعرفي والوجداني والمهاري على مجموعة من الأفراد.

قرارات التوجيه والإرشاد

غالبا ماتستخدم الإختبارات والمقاييس في مساعدة الأفراد في اكتشاف أو اختيار تخصصاتهم أو مهنهم، وفي توجيههم نحو المهن التي اختاروها. والجدير ذكره أن عددا من الأدوات وليس أداة واحدة تستخدم في توجيه وإرشاد الأفراد. فقد يتم استخدام قوائم ميول، واختبارات استعدادات، وقوائم شخصية، واختبارات تحصيل لجعل قرار التوجيه فاعلا وعمليا. وقد يكون قرار الإرشاد متعلقا بالفرد أو بالمؤسسة (المدرسة، مؤسسة اجتماعية أو حكومية، أو مركز مالي أو استثماري).

كما يجدر الإشارة هنا إلى أن قرارات الإختيار والتصنيف والتسكين إدارية، بينما قرارات الإرشاد والتوجيه تشبه قرارات التشخيص والعلاج من حيث كونها تتعلق بالفرد في معظم الأحيان.

قرارات تتعلق بتطوير البرامج

أفضل مثال على هذا القرار ينطبق على المجال التربوي. فـأدوات قيـاس عدة وخاصة التحصيلية منها تستخدم في تطوير وتحسين البرامج التربوية. إذ يمكن تقييم تحصيل الطلبة في مواد دراسية معينة وفي مراحل دراسـية معينة. كما يمكن قياس فعالية التدريس و تصنيف المدارس حسب مـستوى أدائها، ومن ثم توجيه عمليات التطوير التربوي على نحو فعـال. وهنـاك

أدوات شاع استخدامها في بعض البلدان حول تقييم فعالية المدرسة تعرف بالبرامج الإختبارية المدرسية School Testing Programs.

كما يمكن الحكم على مستوى أداء مدرسة ما في ضوء معايير وطنية أو دولية يمكن تحديدها مسبقا (علام، 2000). وعند الحديث عن المدرسة كنظام تربوي، يمتد الحديث إلى تقييم المعلمين وخصائصهم الأكاديمية والتربوية، وطرق التدريس، ونوعية المواد التعليمية والأنشطة المنهجية واللامنهجية في المدرسة. إضافة إلى نوعية علاقة المدرسة بالمجتمع وأولياء الأمور. والجدير بالذكر أن تقويم المدرسة يتم من خلال تقييم مخرجاتها المتمثلة بتحصيل الطلبة وإنجازاتهم من ناحية، ومن خلال دور المدرسة كمؤسسة اجتماعية رائدة، إضافة إلى نوعية المناخ الذي يسود تلك المؤسسة خلال العام الدراسي.

قرارات تتعلق بتحصيل الطلبة

يعتبر تحصيل الطلبة أحد المعايير الرئيسة التي تستخدم في تقييم العملية التعليمية بشكل عام. إذ يرتبط تقييم الطلبة بتقييم التحصيل. وهذا يقود إلى الجوانب التي يجب تقييمها في العملية التدريسية عموما، وفي خصائص المتعلمين ومستوى أدائهم بشكل خاص. ويتبع ذلك ضرورة التمييز بين أنواع الأدوات معيارية المرجع ومحكية المرجع بغرض التوصل إلى تفسير أكثر فاعلية لنتائج الأداء على المهام التربوية والمثيرات العقلية والوجدانية. كما تبرز أهمية تقييم التحصيل من منطلق أن مؤشراته تستخدم في الحكم على صدق كثير من فعاليات تعليمية من ناحية، وأن مصير العديد من الأفراد يتوقف على تقييم مستوى كل منهم في التحصيل في فترة أو مرحلة دراسية معينة.

قرارات تتعلق ببناء وتطوير النظريات النفسية والتربوية

فقد اعتمد بناء أو تطوير النظريات السلوكية على تطبيق أدوات قياس متنوعة، كما اعتمد تطوير نظريات البناء العقلي ونظريات الشخصية في علم النفس على فروض تم التحقق منها بتطبيق عدد من الأدوات ومن ثم تطويرها باستخدام الأساليب الإحصائية المناسبة. وقد ساعد هذا في الوصول إلى التعريف المناسب للمواقف والظواهر في مجالات العلوم الاجتماعية والسلوكية.

ملخص

تتلخص أهم أفكار واستنتاجات ومضامين هذا الفصل في النقاط التالية:

- يمكن تصنيف أدوات القياس حسب عدد من المعايير والشروط. وقد تتداخل هذه التصنيفات. فهناك تصنيف حسب مجالات السلوك الإنساني، أو حسب الهدف، أو طريقة البناء أو طريقة تفسير النتائج، وغيرها.
- يستخدم التقييم معياري المرجع في تفسير درجة الفرد من خلال مقارنتها بدرجات الأفراد الآخرين. بينما يستخدم التقييم محكي المرجع في تفسير أداء الفرد من خلال مقارنة هذا الأداء ببعض المجالات السلوكية المتخصصة بمحك معين.
- من أجل أن تكون الدرجة على الإمتحان ذات معنى كبير، يجب ربطها بالمعايير والمحكات على حد سواء.
- ينبغي على كل اختبار تحصيلي أن يتسم بالصدق سواء استخدم على كان معيارياً أم محكياً.
- إن الإختبار محكي المرجع الخالص، أو بعبارة أدق، الإختبار المجالي المرجع، هو ذاك الإختبار الذي يحتوي على عينة من الأسئلة المأخوذة من المجال المعرفي، بشكل يتيح للفرد أن يقدر نسبة الأسئلة من المجال برمته والذي يعرفه الطالب بالاعتماد على نسبة الإجابات الصحيحة. فقليلة جداً هي الإختبارات التي ينطبق عليها هذا التعريف الضيق.

- تكون الأهداف التي تم انتخاب عينات منها في الإختبار محكي المرجع على الأغلب أكثر ضيقاً في محيط اهتمامها. ولكن تمثيلها من خلال العينة يكون أكثر دقة وعمقاً من تلك الأهداف التي يسعى الإختبار المعياري المرجع إلى تمثيلها.
- تتركز الاستخدامات الرئيسية للقياس محكي المرجع في الوقت الحاضر على اختبارات الإتقان والكفاءة الدنيا والترخيص.
- في اختبارات الإتقان يكون الفرد معنياً باتخاذ قرارات حول ما إذا كان الشخص قد أتقن الهدف بشكل مقنع أو لا.
- قد تكون اختبارات الإتقان أكثر فائدة واستخداما للموضوعات في مستوى المدارس الابتدائية (وبالأخص المراحل الأولى منها).
- يعتمد استخدام الفرد للقياس المعياري المرجع أو المحكي المرجع على نوع القرار المراد اتخاذه.

أسئلة وتمارين

- **ضع دائرة حول رمز الإجابة الصحيحة**

*. يمكن وصف الفارق الرئيس بين اختبارات NRT و CRT، على أساس---.

أ) عدد الفقرات ب) طريقة تفسير النتائج

ج) طريقة تصحيح العلامات د) نوع الفقرات

*. عندما يقال لولي أمر الطالبة إن تحصيل ابنتك في امتحان العلوم أفضل من مستوى أداء 80 بالمائة من طالبات صفها. يكون تصنيف هذا الإمتحان على أنه اختبار---.

أ) مقنن ب) متنبىء

ج) محكي المرجع د) معياري المرجع

*. يصنف مقياس الذكاء عموما على إنه أداة من نوع اختبار ---.

أ) مؤشر ب) فردي ج) ممثل د) عينة

*. يفترض في أداة قياس الأداء العادي أن يظهر المفحوص ---

أ) أفضل أداء يستطيع تقديمه ب) الأداء المرغوب

ج) أدنى أداء ممكن د) الأداء الطبيعي

*. عند الحديث عن الإختبارات، تشير كلمة "موضوعية " إلى ---.

أ) إجراءات التصحيح ب) طريقة تفسير النتائج

ج) نوعية الفقرات المستخدمة د) إجراءات تطبيق الإختبار

*. لتحديد درجة ونوع الصعوبات التي يواجهها مجموعة من الطلبة أو المتدربين، تستخدم أدوات من نوع ---.

أ) قدرة ب) سرعة ج) NRT د) CRT

*. يصنف الإختبار إلى إسقاطي أو محدد البناء حسب درجة تحديد ---.

أ) مجال القياس ب) السلوك المقاس والسلوك المرتقب

ج) مستوى الأداء د) نوع المثير والاستجابة

*. عندما يقال لولي أمر إن ولدكم يستطيع أن يحل 90 % من مسائل القدرة العددية التي تكون اختبار القبول السنوي بنجاح. يكون تصنيف هذا الإختبار على أنه اختبار ---.

أ) مقنن ب) متنبىء

ج) محكي المرجع د) معياري المرجع

*. تستخدم سلالم التقدير في الغالب في تقييم---.

أ) التحصيل ب) الإستعدادات ج) الذكاء د) الاتجاهات

*. الفرق الرئيس بين الإختبارات المقننة وتلك التي من صـنع الـشخص يتركز في--

أ) الموضوعية ب) العينة ج) طريقة البناء د) المعايير

*. تستخدم القوائم في الغالب في تقييم: ---.

أ) النمو العقلي والأكاديمي ب) التوافق الشخصي والميول

ج) الحالات الاجتماعية د) الذكاء والاستعداد الخاص

*. تصنف أداة القياس إلى "أقصى أداء" و"أداء عادي" على أساس ---

أ) طريقة التصحيح ب) نوع الأداء المطلوب في المهمة

ج) شكل الفقرات د) طريقة تفسير النتائج

*. من أهم الوظائف التدريسية لأداة القياس هي ---.

أ) التصنيف والتسكين ب) الإختيار

ج) التشخيص د) التوجيه

المراجع

المراجع العربية

المراجع الأجنبية

9

المراجع العربية

- أبو زينة، فريد ولطفيه، لطفي والخليلي، خليـل. (1981). **الطـرق الإحصائية في التربية والعلوم الإنسانية : نهج حديث**. دار الفرقان للنشر و التوزيع. عمان. الأردن.
- أحمد، محمد عبد السلام. (1960). **القيـاس النفـسي والتربـوي: التعريف بالقياس ومفاهيمه وأدواته، بناء المقـاييس ومميزاتهـا**. مكتبة النهضة المصرية.
- رمضان، زياد. (1981). **مبادىء الإحصاء الوصفي والتطبيقـي والحيوي**. الجامعة الأردنية.
- الزوبعي، عبدالجليل وبكر، محمد و الكنـاني، ابـراهيم. (1981). **الاختبارات والمقاييس النفسية**. مديرية دار الكتب للطباعة والنـشر جامعة الموصل. العراق.
- زيتون، عايش. (1996). **أساسيات الإحصاء الوصفي**. دار عمـار للنشر و التوزيع. عمان . الأردن.
- عدس، عبد الرحمن. (1987). **مبادىء الإحصاء في التربية وعلم النفس**. مكتبة الأقصى. عمان. المملكة الأردنية الهاشمية.
- علام، صلاح الدين. (2000). **القياس والتقويم التربوي والنفسي**. دار الفكر العربي. القاهرة.
- عنبر، أحمـد. (1995). **الأساسـيات فـي الإحـصاء الـسلوكي**. منشورات جامعة دمشق، كلية التربية، دمشق. سوريا.

- عودة، أحمد. (2000). **القياس والتقويم في العملية التدريسية. دار الأمل**. إربد. الأردن.
- الكندري، أحمد. (1995). **علم النفس الاجتماعي والحياة المعاصرة**. مكتبة الفلاح للنشر والتوزيع. العين. الامارات العربية المتحدة.
- النبهان، موسى. (2001). **أساسيات الإحصاء في التربية والعلوم الإنسانية والاجتماعية**. مكتبة الفلاح للنشر والتوزيع. العين. الامارات العربية المتحدة.

المراجع الأجنبية

- Allen, M. J and Yen, W. M. (1979). **Introduction to Measurement Theory**. Monterey, CA: Brooks/Cole.
- Anastasi, A. (1976). **Psychological Testing**. 4th Ed. Macmillan Publishers.
- Baker, E. L. (1987). Issues in Policy , assessment, and equity. **In Proceedings of the National Proficiency Student Issues: Vol. 1 and 2: Focus on Evaluation and Measurement,** Washington, DC.
- Bartz, A. E. (1988). **Basic Statistical Concepts**. 3rd Edition McMillan Publication Company.
- Besag, F. P. and Besag, P. L. (1985**). Statistics for the Helping Professions**. Sage Publications.
- Beuchert, A. K. and Mendoza, J. L. (1979). A Monte Carlo comparison of ten item discrimination indices. **Journal of Educational Measurement, 16**, 109-118.
- Blommers, P. J. and Forsyth, R. A. (1977). **Elementary Statistical Methods in Psychology and Education**. 2nd Edition. Houghton Mifflin Company.
- Bloom, B. S.(1968). Learning for Mastery. **Evaluation Comment, 1**, 1-12.

- Bloom, B. S. (1965). **Taxonomy of Educational Objectives: Cognitive Domain**. New York: David Mckay.
- Brennan, R. (1972). **Elements of Generalizability Theory**. Iowa: ACT Publication.
- Brown, F. (1976). **Principles of Educational and Psychological Testing**. 2nd Ed. Holt, Rinehart and Winston.
- Brown, F. (1981). **Measuring Classroom Achievement**. New York: Holt, Rinehart and Winston.
- Buros, O. K. (Ed.). (1995). **The Mental Measurement Yearbook**. Buros Institute of Mental Measurements.
- Campbell, D. T. and Fiske, D. W. (1959). Convergent and Discriminant validation by the multitrait-multimethod matrix. **Psychological Bulletin, 56**, 81-105.
- Chase, Clinton, I. (1978). **Measurement for Educational Evaluation**. (2nd. Ed.). Addison-Wesley Publishing Company.
- Cook, D., Grawen, A. and Clarke, G.(1999). **Basic Statistical Computing.** Edward Arnold Publishers. London.
- Cox, R. and Vargas, J. A. (1966). **A Comparison of Item Selection Techniques for Norm-Referenced and**

Criterion-Referenced Tests. Paper presented at the annual meeting of the NCME. Chicago.

- Crocker, L. and Algina, J. (1986). **Introduction to Classical & Modern Test Theory**. Holt, Rinehart and Winston.
- Cronbach, L. (1980). **Toward Reform of Program Evaluation**. Calif. Jossey-Bass Publishers.
- Cronbach, L. J. and Meehl, P. E. (1955). Construct validity in psychological tests. **Psychological Bulletin, 52**, 281-302.
- Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.). **Educational Measurement** (2^{nd} Ed.). Washington D.C.: American Council on Education.
- Cronbach, L. J. and Others. (1972). **The Dependability of Behavioral Measurements**. New York: John Wiley.
- Cureton, E. E. (1958). The definition and estimation of test reliability. **Educational Psychological Measurement, 18,** 715-738.
- Dunteman, G. H. (1984). **Introduction to Linear Models**. Sage Publications.
- Ebel, R. (1979). **Essentials of Educational Measurement**. (3^{rd}. Ed.) N.J.:Englewood Cliffs.

- Englehart, M. D. (1965). A comparison of several item discrimination indices. **Journal of Educational Measurement, 2**, 69-76.
- Ferguson, G. A. (1966). **Statistical Analysis in Psychology and Education**. 2nd Edition. McGraw-Hill.
- Flanagan, J. C. (1954). The critical incident technique. **Psychological Bulletin, 51**, 327-358.
- Feldt, L. S. and Brennan, R. L. (1989). Reliability. In R. L., Linn (Ed.), **Educational Measurement** (3rd ed.). Englewood Cliffs, NJ: Merrill/Prentice Hall.
- Gay, L.R.; Mills, G. and Airaisan, P. (2008). **Educational Research: Competencies for Analysis and Application**. 6th. Edition. Prentice-Hall Inc.
- Glaser, R .(1963). Instructional technology and the measurement of learning outcomes. **American Psychologist, 18**, 519-521.
- Glaser, R. and Nitko, A. (1971). Measurement in learning and instruction. In R. L. Thorndike (Ed.). **Educational measurement** (2nd ed., pp. 625-670). Washington, DC: American Council on Education.
- Glass, G. and Hopkins, K. (1996). **Statistical Methods in Education and Psychology**. Prentice - Hall, Inc.
- Glassnapp , D. R. and Poggio, J. P. (1985). **Essentials of Statistical analysis for the Behavioral Sciences**. Charles, E. Merrill Publishing Company.

- Gravetter, F. J. and Wallnau, L. B. (1991). **Essentials of Statistics for the Behavioral Sciences**. West Publishing Company.
- Gronlund, N. E. (1985). **Preparing Criterion Referenced Tests for Classroom Instruction**. New York: Macmillan.
- Haladyna, T. (1980). On the Edumetric Psychometric Dimensions of Tests. **American Psychologist, 30**.
- Hambleton, R. K. (1980). **Test score validity and standard setting methods**. In R. A. Berk (Ed.). Criterion-referenced measurement: The state of the art. Baltimore: John Hopkins University Press.
- Hambleton, R. K. and Novick, R. (1973). Towards an Integration of Theory and Method for Criterion-Referenced Tests. **Journal of Educational Measurement, 10**, 159-170.
- Harris, C. W., et al. (eds.). (1974). **Problems in Criterion-Referenced Measurement. CSE Monograph Series in Evaluation, No. 3**, Center for the Study of Evaluation, University of California, Los Angles.
- Hastings, N. A. J. and Peacock, J. B. (1975). **Statistical Distributions**. John Wiley & Sons.
- Hieman, G. w. (1992). **Basic Statistics for the Behavioral Sciences**. Houghton Mifflin Company.

- Hills, J. (1981). **Measurement and Evaluation in the Classroom**. Ohio: A Bell and Howell Company.
- Hinkle, D., Wiersma, W. and Jurs, S. G. (1979). **Applied Statistics for the Behavioral Sciences**. Rand McNally College Publishing Company .
- Howell, D.C. (1999). **Fundamental Statistics for the Behavioral Sciences.** 4th Ed. Brooks/ Cole Publishing Company.
- Hughes, H. H. and Trimble, W. E. (1965). The use of complex alternatives in multiple-choice items. **Evaluational and Psychological Measurement, 25**, 117-126.
- Isaac, S. and Michael, W.B. (1984). **Handbook in Research and Evaluation**. 4th. Edition, Edits Publishers.
- Jackson, D. (1970). **Problems in Human Assessment**. N. Y: McGraw-hill.
- Kagan, J. and Mc Call, R. B. (1988). **Fundamental Statistics for Behavioral Sciences**. 4th. Edition. Harcourt Brace Jovanovich Publishers.
- Kane, M. T. (1982). A sampling model for validity. **Applied Psychological Measurement, 6**, 125-160.
- Katz, M. (1958). Selecting an achievement test: Principles and procedures. In V. H. Noll, D. P. Scannell, and R. P. Noll (Eds.). **Introductory readings in**

educational measurement. Boston: Houghton Mifflin, 1972.

- Kellogg, C. E. and Morton, N. W. (1999). **Beta III Manual**. The Psychological Corporation.
- Kelly, E. F. (1975). **Assessment of Human Characteristics**. Calif: Brooks/ Cole Publishers.
- King, J. A. and Evans, K. M. (1991). Can we achieve outcome-based education?. **Educational Leadership, 49**(2). 73-75.
- Klein, S. P. and Kosecoff, J. P. (1975). **Determining how well a test measures your objectives.** (CSE Report No. 94). Los Angles: Center for the study of Evaluation, University of California.
- Kleinbaum, D. G., Kupper, L. L. , Muller, K. E. (1988). **Applied Regression Analysis and other Multivariate Methods**. 2nd Edition. PWS-Kent Publication Company.
- Krathwogl, D. R., Bloom, B. S. and Masia, B. (1964). **Taxonomy of educational objectives. handbook 11: The affective domain**, New York: McKay.
- Lindquist, E. F. (1936). **The theory of test construction**. In H. E. Hawkes, E . F. Lindquist, and C. Mann (Eds.). The construction and use of achievement examinations, Boston: Houghton Mifflin.

- Linn, R. L. (1980). Issues of Validity of Criterion-Referenced Measurement. **Applied Psychological Measurement, 4**, 4, 547-561.
- Loevinger, J. (1979). Construct validity of the Sentence Completion Test of Ego Development. **Applied Psychological Measurement, 3**, 281-311.
- Lyman, Howard, B. (1991). **Test scores and what they mean**. (5th. Ed.). prentice Hall, Englewood Cliffs, NJ 07632.
- Magnusson, D. (1967). **Test Theory**. Translated from Swedish by Hunter Mabon. Addison-Wesley Publishing Company.
- Marascuilo, L. A. and Serlin, R. C. (1988). **Statistical Methods for the Social and Behavioral Sciences**. W. H. Freeman and Company.
- McClung, M. S. (1978). **Developing proficiency program in California public schools: Some legal implications and a suggested implementations schedule**. Sacramento: California State Development of Education.
- Mehrens, W. and Lehman, L. (1991). **Measurement and Evaluation in Education and Psychology**. (4th Ed.). Fort Worth: Holt, Rinehart & Winston.

- Messick, S. (1989). Validity. In R. L., Linn. **Educational Measurement**. (3rd. Ed.). (pp. 13-103). American Council of Measurement in Education.
- Millman, J. (1974). **Criterion-Referenced Measurement**. In W. J. Popham (ed.), Evaluation in Education: Current Application. Berkley, Calif.: McCutchan, Chapter 6.
- Murray, H. A. (1938). **Exploration In Personality**. NewYork: Oxford University Press.
- Nitko, A. J. (2005). **Educational Assessment of Students**. 3rd . Ed. Merrill Prentice Hall.
- Nitko, A. J. (1983). **Educational Tests And Measurement: An Introduction**. Harcourt Brace Jovanovich, Inc.
- Nitko, A. J. and Hsu, T. C. (1974). Using domain referenced tests for students placement , diagnosis, and attainment in system of adaptive individualized instruction. In W. Hively (Ed.), **Domain-referenced testing**. Englewood. Cliffs, NJ: Educational Technology Publications.
- Nunnally, J. C. (1978). **Psychometric Theory**. McGraw-Hill, Inc.
- Osterlind, S. (1983). **Constructing Test Items**. MA. : Kluwer Academic Publishers.

- Popham, W. J. (1981). **Modern Educational Measurement**. Englewood Cliffs, N. J. : Prentice-Hall.
- Popham, W. J. (1969). **Educational Evaluation**. Englewood Cliffs, NJ: Prentice Hall.
- Sax, G .(1981). **Principles of educational and psychological measurement and evaluation** (2nd Ed.). Belmont, Calif.:Wadsworth.
- Scriven, M. (1967). **The Methodology of Evaluation**. (AERA Monograph Series of Curriculum \evaluation, No. 1). Chicago: Rand McNally.
- Shavelson, R. and Stern, N .(1981). Generalizability theory: 1973-1980. **British Journal of Mathematical and Statistical Psychology**.
- Shmidt, F. L.; Hunter, J. E. and Urry, V. W. (1976). Statistical power in criterion-related validity studies. **Journal of Applied Psychology, 61**, 473-485.
- Shoemaker, D. M. (1975). Toward a framework for achievement testing. **Review of Educational Research, 45**, 127-148.
- Spatz, C. and Johnston, J. O. (1989). **Basic Statistics: Tales of Distributions**. 4th. Edition. Brooks / Cole Publishing Company.
- Stanley, J. C. (1981). Reliability in R. L. Thorndike (Ed.). **Educational measurement (2nd Ed.)**. (pp. 359-

442). Washington, D.C.: American Council on Education.

- Stevens, S. S. (1946). On the theory of scales of measurement. **Science. 103**, 667-680.
- Stiggins, R. J. and Conklin, N. F., and Associates. (1992). **In teachers' bands: Investigating the practice of classroom assessment**. Albany, NY: SUNY Press.
- Swaminathan, H., R. K. Hambleton, and J. Algina. (1975). A Bayesian Decision-Theoretic Procedure for Use with Criterion-Referenced Tests. **Journal of Educational Measurement, 12**, 87-98.
- Thorndike, R. L. and Hagen, E. (1986). **Measurement and evaluation in psychology and education**. New York: John Wiley.
- Trice, Ashton, D. (2000). **A Handbook of classroom assessment**. Addison Wesley Longman, Inc,
- Webb, E. J., Campbell, D. T., Schwarts, R. D. and Sechrest, L. (1981). **Unobtrusive measures: Non-reactive research in the social sciences**. Chicago: Rand McNally.
- Weinberg, Sharon and Goldberg. (1990). **Statistics for the Behavioral Sciences.** Cambridge University Press.
- Witte, R. S. (1993). **Statistics** .4th edition. Harcourt Brace Jovanovich College Publisher.

- Worthen, B. R. and Sanders, J. R. (1987). **Educational Evaluation: Alternative Approaches and Practice Guidelines**. Longman.
- Yalow, E. S. and Popham, W. J. (1983). Content validity at the crossroads. **Educational Researcher, 12**(8), 10-14.

كشاف مصطلحات ومفاهيم القياس والتقييم والتقويم

A Glossary of Measurement, Assessment and Evaluation Terms

10

القياس Measurement

هو التعبير الكمي عن الخصائص أو السمات أو الأحداث بناءا على قواعد وقوانين محددة. والمعروف أن التعبير الكمي عن السمة يفيد في تقدير شدة وجودها من ناحية، ويساعد في إجراء المقارنات بشكل أكثر دقة من ناحية أخرى.

الاختبار Test

إجراء منظم لقياس عينة من السلوك، أو هو إجراء منظم لملاحظة ووصف سمة أو أكثر من سمات الفرد بالاستعانة بمقياس أو نظام تصنيف معين. ويحدث أحيانا خلط بين مفهومي الاختبار والامتحان Examination. إذ أن الاختبار أشمل من الامتحان، ويتعلق الأخير بقياس الأداء في المجالات التحصيلية. من ناحية أخرى، يحدث التباس بين مفهومي الاختبار والمقياس Scale . فإذا كان الاختبار مرتبطا أكثر بقياس الأداء في المجالات المعرفية (الذكاء والاستعدادات والتحصيل)، فإن المقياس أداة قياس للجوانب المعرفية وكذلك الوجدانية أو تلك التي يتم قياسها على سلم يتألف من عدد من الدرجات (3)، أو (5)، أو (7). كذلك، ليس لفقرات المقياس إجابة نموذجية، بل ربما يشير الأداء على فقرة المقياس إلى مدى انطباق الحالة أو شدة انطباقها على المستجيب.

القدرة Ability

هي كفاية الفرد في مجال محدد، وهي خاصية من خصائص الفرد. وغالبا ما يتم الخلط بين القدرة والاستعداد. فبينما ترتبط القدرة بالحالة القائمة للشخص، يرتبط الاستعداد بالواقع المستقبلي أو التنبؤي له.

السمة Trait

مفهوم يستخدم في وصف سلوك الأفراد. وهو عبارة عن سلوك أو مجموعة من السلوكات التي يميل الفرد إلى إظهارها. فالسمة ليست صفة مفردة، وليست شيئا ملموسا بل هي مفهوم مجرد. وهناك سمات عقلية وشخصية ووجدانية. كذلك هناك سمات تتعلق بالجانب الجسمي كالخصائص الجسمية كشكل الصدر، والطول، والوزن، وحجم الكفين والقدمين، والخصائص الفسيولوجية المتعلقة بالسلوك الداخلي لأعضاء الجسم الحيوية كضغط الدم ونشاط الغدد. وبشكل عام تتمتع السمات العقلية بثبات أكبر نسبيا من السمات الوجدانية.

التحصيل Achievement

المستوى الذي تعلمه الفرد للقيام بالأداء على مهارة معينة. وعادة ما يرتبط التحصيل بمجمل المعلومات والمهارات والتمارين والأفكار التي تم اكتسابها خلال صف أو مرحلة دراسية معينة.

الاستعداد Aptitude

تجميع من الخصائص الموروثة والمكتسبة التي تشير إلى قدرة الفرد أو قابليته لأداء عمل معين أو تعلم مجال ما في المدى الزمني القادم إذا حصل على التدريب المناسب. ومن بين الاستعدادات: الدراسي، والميكانيكي، والكتابي، والموسيقي والاستعداد للأداء في الرياضيات.

الكفاية Competency

هي شرط تحقيق مستوى (معيار) معين من الأداء (التحصيل مثلا) لدى الفرد.

الاختبار الرسمي (المقنن) Formal (Standardized) Test

لاختبار الذي يتم بناؤه وتطويره من قبل فريق من المختصين ولصالح مؤسسة معينة. كما يتم تطبيقه على عينة كبيرة نسبيا من الأفراد. ويتم تطوير دليل Manual له يتضمن تعليمات خاصة بالتطبيق والتصحيح ومعلومات عن مؤشرات الصدق والثبات وخصائص الفقرات وتفسير النتائج عليه في ضوء معايير Norms خاصة يتم اشتقاقها. وتعرف الاختبارات المقننة بالمنشورة. ومن أمثلتها اختبارات الذكاء وقوائم الاتجاهات ومقاييس الشخصية واختبارات التحصيل الوطنية مثل امتحان الثانوية العامة في معظم البلدان.

الاختبار غير الرسمي Informal Test

وتعرف أحيانا بالاختبارات من صنع الشخص Man-made tests وهي الاختبارات التي يتم بناؤها أو تطويرها من قبل معلم المدرسة أو أستاذ الجامعة أو مسؤول البرنامج التدريبي، بغرض قياس مستوى تحصيل الطلبة أو كفاءة المتدربين. وغالبا ما تعرف هذه الاختبارات بالامتحانات. ويبين الجدول التالي مقارنة بين الاختبارين (الرسمي وغير الرسمي)

التقنين Standardization

عملية تطوير مقاييس واختبارات (مقننة أو في طور الإعداد) من أجل تحسين ملاءمة استخدامها لصالح مجتمع طور المقياس لخدمته، أو لصالح مجتمع غير الذي بني الاختبار من أجله. ويتطلب التقنين تجريب فقرات المقياس، وتحليل الأداء عليها. كما يتضمن التقنين تقييم درجة صدق الاختبار وتقدير مستوى ثباته. ويشمل كذلك التقنين اشتقاق أو تطوير معايير خاصة لاستخدام المقياس أو الاختبار **(Lyman, 1991)**.

من ناحية أخرى، يؤدي التقنين إلى تطوير موضوعية المقياس، ويحسن مستوى تفسير درجات الأفراد الممتحنين عليه **(Nitko, 1983)**. وفي بعض الأحيان، يتم ترجمة فقرات المقياس وتعليماته إلى لغة المجتمع الذي سيتم استخدام المقياس فيه، كما أنه من الممكن أن يتم تطوير أو تعديل أو استبدال أو حذف بعض فقرات المقياس التي لا تتلاءم أو قد تتعارض مع ثقافة مجتمع معين.

وعندما يصبح الاختبار مقننا في بيئة اجتماعية معينة، فهذا يعني أنه يجوز تطبيق ذلك الاختبار على عينات أخرى من تلك البيئة باتباع التعليمات نفسها، وطرق التصحيح نفسها، واستخدام نفس الفترة الزمنية. ويتم تفسير أداء الفرد على ذلك الاختبار في ضوء بيانات معيارية مشتقة من تلك البيئة الاجتماعية.

هذا بالنسبة لأدوات القياس النفسية والاجتماعية أو تلك المتعلقة بقياس السلوك الإنساني، نظرا لاختلاف طبيعة ذلك السلوك وما يؤثر فيه من مؤثرات كمية ونوعية. أما بالنسبة للمقاييس في المجالات الطبيعية، فالصورة تختلف، من حيث أن تلك الأدوات لا تختلف خصائصها باختلاف البيئة أو الظروف المحيطة عموما، كما في أداة قياس الطول (المتر ، أو القدم).

البطارية Battery

مجموعة مختارة بعناية من الاختبارات المترابطة التي تقيس سمة أو قدرة واحدة أو أكثر، يتم عرضها على مجتمع معين من الأفراد، وعادة ما يتم تقييم مستوى الأداء على أحد اختباراتها، أو على عدد منها، أو جميعها. ومن بطاريات الاختبار المشهورة بطارية Graduate Register Examination (GRE) وبطارية الاستعدادات الفارقة Differential

أو بطارية التحصيل الدراسية Aptitude Test (DAT) Scholastic Achievement Test (SAT).

الفقرة Item

هي السؤال، أو العبارة، أو التمرين، أو المهمة، أو المسألة التي يتضمنها الاختبار أو المقياس أو أداة التقييم. وتصاغ الفقرات بأكثر من نمط، فمنها الفقرات الانتقائية والفقرات الصياغية.

الفقرة الانتقائية Selective Item

ذلك النوع من الفقرات التي يطلب فيها من المفحوص أن يختار الإجابة الصحيحة من بين عدد من البدائل والعبارات المعطاة. ومن أنواع هذه الفقرات: فقرات الاختيار من متعدد Multiple Choice Items، وفقرات الصواب والخطأ True-False Items، وفقرات المزاوجة Matching Items.

الفقرة الصياغية Supply Item

ذلك النوع من الفقرات التي يطلب فيها من المفحوص أن يكتب أو يصيغ (ينشيء) الإجابة الصحيحة على السؤال، وتعرف أحيانا بالفقرة الإنشائية. ومن أنواع هذه الفقرات: فقرات الإكمال Completion وفقرات الاجابة القصيرة Short answer والفقرات المقالية Essay وفقرات حل المسألة Problem solving أو الرسم Drawing وعمل النماذج Designing وغيرها.

القائمة Inventory

هي دليل أو قائمة من المفردات أو العبارات تهدف إلى تقييم مدى وجود اتجاهات أو اهتمامات أو سلوكات محددة لدى مجموعة من الأفراد. غالبا ما تعرف كأدوات ورقة–وقلم تستخدم في تقويم الشخصية أو

الاتجاهات أو الميول. وتعرف القوائم أحيانا بأنها أدوات التقرير الذاتي Self- Report Tools. ويندر استخدام القوائم كأدوات قياس تحصيل أو مهارات أو قدرات أو استعدادات.

أدوات إسقاطية Projective Tools

هي مجموعة الأدوات التي لا يكون أي من المثير أو الاستجابة محددا كمقياس البقع الحبرية (الرورشاخ)، واختبار تفهم الموضوع (TAT) وبعض الاختبارات والمقاييس النفسية الأخرى. وتستخدم هنا عبارات معينة من أجل المساعدة في الوقوف على مستوى الطلبة في مجالات محددة كمستواهم في التعبير والكتابة. كما قد تستخدم الأدوات الإسقاطية في استقصاء نوعية وحجم المشكلات التي يواجهها الأفراد في دراستهم أو في حياتهم الأسرية وغيرها

أدوات محددة البناء Structured Tools

هي مجموعة الأدوات التي يكون فيها كل من المثير والاستجابة محددا، كما في اختبارات التحصيل والقدرات والاستعدادات. أي أن لها أسئلة محددة وإجابات محددة.

أدوات ذات الأداء الأقصى Maximal Performance Tools

هي مجموعة الأدوات التي يطلب فيها من الممتحن أن يظهر أفضل ما لديه من قدرات أو معلومات عند الإجابة عليها، كما في اختبارات المهارات والتحصيل والقدرات والاستعداد. إذ تتضمن هذه الأدوات إجابة صحيحة أو نموذجية. لذلك فإن من الممكن تشجيع الطلبة على الدراسة والتعمق لبلوغ درجات تحصيل أفضل على هذا النوع من الاختبارات.

أدوات ذات الأداء العادي Typical Performance Tools

هي مجموعة الأدوات التي يطلب من الممتحن أن يبذل أداء عاديا كما في قوائم الميول والاتجاهات والشخصية وغيرها. إذ لا تتضمن هذه الأدوات إجابة صحيحة أو نموذجية، بل تعكس سلوك المتعلم في الأوضاع العادية. ولا حاجة إلى تشجيع الطلبة على الدراسة للاستجابة عليها، بل إلى توخي الدقة والموضوعية في الإجابة.

الاختبارات الفردية Individualized Tests

هي مجموعة الاختبارات التي يمكن أن تطبق بصورة فردية (أي تطبق على كل مفحوص لوحده). كما في اختبارات الشخصية والامتحانات الشفوية والفحص الطبي وبعض اختبارات الذكاء مثل (ستانفورد - بينيه) وويكسلر لذكاء الأطفال وأطفال ما قبل المدرسة، والراشدين، إضافة إلى اختبارات القدرات الفنية وبعض اختبارات المهارات. ويشيع استخدام مثل تلك الأدوات في الأغراض الاكلينيكية. وتتيح فرصة التفاعل مع المستجيب، أو التعرف على ردود أفعال المستجيبين أثناء التطبيق.

الاختبارات الجمعية Grouped Tests

هي مجموعة الاختبارات التي يتم تطبيقها على مجموعات من الأفراد، كما في الاختبارات التحصيلية الصفية والاختبارات المقننة واختبارات القدرات العقلية كما في اختبار كاليفورنيا للنضج العقلي، واختبار بيتا للذكاء غير اللفظي، واختبار أوتيس لينون للقدرة العقلية العامة، واختبار كالمان وأندرسون للذكاء.

اختبارات القوة (القدرة) Power Tests

مجموعة الاختبارات التي يكون الزمن المخصص للإجابة عليها مفتوحا نسبيا كما في اختبارات التحصيل والقدرات والاستعدادات. وهنا لا

تكون السرعة في الاستجابة مسألة حاسمة، بل هي أمرا ثانويا. إذ أن الموضوع الهام هنا هو قدرة المفحوص على الاستجابة، وليس سرعته في إنجازها. وخير مثال على ذلك الاختبار المتعلق بفحص طبيب جراح متقدم لشغل وظيفة في مستشفى، فهذا في الغالب اختبار قدرة، حيث أن إتقان إجراء العلمية الجراحية أكثر أهمية من السرعة في إجرائها.

اختبارات السرعة Speeded Tests

مجموعة الاختبارات التي يكون الزمن المخصص للإجابة عليها محددا نسبيا، كما في الاختبارات المتعلقة ببعض المهارات مثل سباقات الجري والسباحة والقفز أو حتى الطباعة. وهنا تعد السرعة في الاستجابة مسألة حاسمة، إذ أن الموضوع المهم هنا هو ارتباط أداء المفحوص بسرعة الاستجابة أو زمن الاستجابة المحدد. وتكون فقرات اختبارات السرعة عموما سهلة، إذ يستطيع المفحوص الإجابة على كافة فقرات الاختبار لو أعطي الوقت الكافي.

اختبارات موضوعية Objective Tests

هي أدوات القياس التي لا تتأثر درجات الممتحنين عليها بشخصية المصحح، أو وجهة نظره، أو خلفيته الاجتماعية أو السياسية. إذ يتم إجراء عملية التصحيح باستخدام مفتاح مثقب لهذا الغرض، أو تتم العملية آليا، أو باستخدام الحاسوب. وهذا إنما ينطبق بشكل عام على تصحيح الاختبارات ذات الفقرات الانتقائية.

اختبارات ذاتية Subjective Tests

هي الأدوات التي تتأثر درجات الممتحنين عليها بشخصية المصحح أو وجهة نظره أو خلفيته الاجتماعية أو السياسية. إذ يتم إجراء عملية التصحيح من قبل مصححين ربما يعرفون الممتحنين بشكل أو بآخر. وهذا

يعتمد على فهم وتقدير هؤلاء الأشخاص لمستوى أداء الممتحن. ويبدو ذلك واضحا أكثر في حالة الاختبارات ذات الإجابة المصوغة وخصوصا تلك الإنشائية ذات النهاية المفتوحة، كاختبارات التعبير الأدبي.

مقاييس لفظية Verbal Scales

هي تلك الأدوات التي يتم استخدام اللغة فيها بشكل رئيس في التعليمات والفقرات ذاتها. وهذا إنما يتطلب قدرة الممتحن على القراءة والكتابة. وتشيع هذه الأدوات في المجالات المعرفية والوجدانية، وتقل نسبيا في مجال المهارات الحركية، والقدرات الحسابية، والمهارات الرياضية.

مقاييس غير لفظية Non-Verbal Scales

أو ما تدعى أحيانا بمقاييس الأداء العملي، وهي تلك الأدوات التي لا يتم استخدام اللغة فيها بشكل رئيس سواء في التعليمات أو في محتوى الفقرات ومدلولها. وهذه لا تتطلب كثيرا قدرة الممتحن على القراءة والكتابة، كما في الأدوات التي تستخدم مع الأطفال المعاقين أو الأطفال الصغار أو الأميين.

إنها فقرات تتطلب إكمال الصور وترتيبها، وتجميع الأشياء، وتصميم المكعبات، وتتبع المتاهات. ولا يخفى أن هناك بعض فقرات اختبارات الذكاء الفردية التي تتضمن تقييم الاستعداد المكاني والحركي وبعض الجوانب النفسية والحركية. من ناحية ثانية، تعرف هذه المقاييس بأنها مقاييس متحررة ثقافيا Culture-Fair Tests. ومن أمثلتها اختبار بيتا Beta III للذكاء غير اللفظي، واختبار رافن للمصفوفات المتتابعة.

التقويم Evaluation

عملية قياس ومقارنة وتقدير وتثمين نوعية أعمال الأفراد أو المؤسسات أو البرامج أو المشاريع واتخاذ القرار المناسب بشأن كل منها.

وهناك أكثر من تصنيف لأنواع التقويم، أشهرها تصنيف التقويم في (1) تقويم بنائي (تكويني) و (2) تقويم تجميعي (ختامي).

التقويم البنائي أو التكويني Formative Evaluation

عملية تشخيصية تصحيحية مستمرة، تهدف إلى تقويم تحصيل أو أداء مجموعة من الأفراد وتحديد احتياجاتها، وتشخيص جوانب القوة والضعف في أداء الفرد أو المؤسسة. كما يهدف (تربويا) إلى ضمان إحداث التعلم المطلوب مع توفير تغذية استرجاعية Feedback لمعاودة التعليم والتقويم بقصد تحسين مردود عملية التعلم والتعليم وبناء الشخصية.

التقويم الختامي (التجميعي) Summative Evaluation

عملية تلخيص طولية تقوم على التقدم عبر الزمن وترمي إلى إبراز أهم النتائج المتسقة المتكاملة والمبنية على مجمل التقويم البنائي (التكويني). كما يهدف إلى تقويم مستوى تمكن أو إتقان مجموعة من الأفراد وتقويم مكانة الفرد من أجل تسكينه في برنامج معين أو تحديد المستوى الذي حققه في مرحلة من مراحل تدريبية أو تعليمية متسلسلة من أجل اتخاذ قرار الانتقال إلى المرحلة التالية ، وكذلك تقويم الفاعلية النسبية لبرامج أو مناهج أو حتى مشاريع معينة.

التقويم الكمي Quantitative Evaluation

هو التقويم الذي يعتمد على البيانات الكمية أو التي يمكن أن يعبر عنها بالأرقام. كالبيانات التي نحصل عليها من الاختبارات أو الاستبانات وبعض أدوات القياس الأخرى. وهنا يتم التعامل مع هذه البيانات من خلال مؤشرات إحصائية معروفة كمؤشرات النزعة المركزية والتشتت والمئينات.

التقويم النوعي Qualitative Evaluation

التقويم الذي يعتمد على البيانات التي نجمعها من الواقع الميداني من خلال الملاحظة والمقابلة وملفات الانجاز وتقييم الرفاق وأولياء الأمور وغيرها. وبطبيعة الحال، يعتمد قرار التقويم عموما على نوعي البيانات ولكن ربما بدرجات تركيز تختلف حسب موضوع التقويم ونوع الأسئلة المنوي الإجابة عليها.

تقويم البرامج Programs Evaluation

التقويم الذي يختص بالحكم على كفاءة برنامج معين من خلال الحكم على ثلاث مكونات هي **المدخلات والعمليات والنتاجات**. ويتركز تقويم المدخلات Inputs Evaluation على عناصر البرنامج المادية والبشرية. ففي البرنامج التربوي، تعد خصائص الطلبة، والإمكانات المتوفرة في المؤسسة التربوية من تسهيلات الطباعة والتصوير ومؤهلات العاملين في البرنامج ومستوى صياغة فقرات الاختبار وغيرها بالمدخلات.

أما **تقويم العمليات** Processes Evaluation ، فهو مجموعة الإجراءات التي توجه نحو تقويم البرنامج أثناء عمليات التنفيذ، كما في طرق التدريس وطبيعة التفاعل الصفي والتوقيت الزمني لفعاليات البرنامج وخطواته، ويتضمن ذلك إجراءات سير الامتحانات وطبيعة العلاقات بين الطلبة والمعلمين وعناصر الهيئة الإدارية. ويكون **تقويم النتاجات** Products Evaluation عن طريق اتخاذ قرارات تتعلق بمجمل النواتج التي ستنتج كخلاصة للعمليات التي حدثت في البرنامج. ومثال ذلك مجمل أعمال الطلبة الكتابية، والمهاراتية كالرسم وكتابة التقارير وإجراء التجارب، وتسجيل المشاهدات ، وتحضير الشرائح، وتطوير النماذج وغيرها.

التقويم الشمولي أو المكبر Macro Evaluation

نوع من التقويم الذي يتطلب مشاركة كافة المعنيين في البرنامج كفريق ويستغرق وقتا طويلا نسبيا ويهدف إلى تقويم البرنامج بمكوناته معا دون إغفال دور أي من هذه الجهات كالمعلمين، والمديرين، والمرشدين، وجهات التمويل، وأولياء الأمور، ومصممو المناهج الدراسية، ومؤلفو الكتب الدراسية، وفنيو المختبرات، ومنظرو الفلسفات والسياسات التربوية.

التقويم الجزئي Micro Evaluation

نوع من التقويم الذي يهتم بدراسة مكون واحد من مكونات البرنامج، كتقويم مدى فعالية استخدام المختبرات المدرسية في مجموعة مدارس ثانوية. ويختلف التقويم الشمولي عن التقويم الجزئي من حيث مقدار الجهد المطلوب بذله في عمليات كل منهما (Worthen and Sanders, 1987). وجدير بالذكر بأن الشمولية تعد نسبية، وذلك عندما يصار إلى تجزئة البرنامج الجزئي إلى أجزاء أصغر.

تقويم داخلي Internal Evaluation

نوع من التقويم الذي يقوم به مجموعة من الأشخاص من داخل البرنامج موضع التقويم. كأن تقوم هيئة خاصة بتقويم فعالية مشروع أو برنامج معين، يتم تشكيلها كمؤسسة مساندة للتطوير.

تقويم خارجي External Evaluation

نوع من التقويم الذي يقوم به أناس من خارج البرنامج أو الدائرة موضع التقويم. كأن تتولى مؤسسة متخصصة القيام بتقويم برنامج معين أو مؤسسة معينة عن طريق التعاقد معها لهذا الغرض.

الغاية Aim or Goal

عبارات عامة تصف التوقعات المنوي تحقيقها جراء عملية التدريس. والغاية أكثر عمومية من الهدف، ويحتاج تحققها زمنا طويلا نسبيا (مرحلة دراسية أو أكثر)، ولا يسهل تحققها خلال الدرس اليومي. ولأن الغايات تنبثق من القاعدة الفلسفية للمنهاج، فهي تغذي التوجه العام وتساعد في تطويره.

الهدف Objective

هو النتاج التعلمي المتوقع من المتعلم اكتسابه خلال فترة زمنية قصير نسبيا (حصة دراسية مثلا). ومن المفروض أن يصاغ الهدف صياغة سلوكية ليتمكن المعلم من قياسه بفاعلية، وهذا يعرف بالهدف السلوكي Behavioral Objective وهناك الهدف التربوي الذي تصل عموميته إلى مستوى الغاية. وهناك الهدف الضمني Implicit Objective الذي يتوسط بين الهدف السلوكي والهدف التربوي.

المخرج Outcome

وصف للإمكانات والمعارف التي من المفروض أن يظهرها المتعلم نتيجة لخبراته التربوية. والمخرج هدف تربوي معرف إجرائيا، عادة ما يكون نشاطا واضحا، أو نتاجا، أو أداء يمكن قياسه. والمخرج هو الناتج النهائي للعملية التدريسية أو التدريبية، وقد يكون ملاحظا أو على هيئة تغيرات داخلية في المتعلم. وبينما يتم اشتقاق الهدف من المحتوى أو مقررات الكتاب الدراسي، يعتمد المخرج على نوع التغيرات المطلوب تحققها في المتعلم (King and Evans, 1991). وعلى الرغم من أن معنى كل من الغاية والهدف والمخرج يبدو متقاربا في المعنى ويحل كل منها مكان الآخر، إلا أن البعض يرى فيهما اختلافا بينا وتمييزا واضحا.

المحك Criterion

مجموعة (مجال) من المعارف والمهارات المحددة تحديدا جيدا بحيث يمكن نتيجة لمقارنة أداء الفرد في الاختبار بهذا المجال أن نعرف ما يستطيع الفرد أن يؤديه وما الذي لا يستطيع. والمقارنة هنا تكون في ضوء فقرات الاختبار ذاتها. فكل فقرة يشملها الاختبار تمثل معلومة أو مهارة ينبغي أن يكتسبها الفرد لكي ينجح في أداء عمل معين. حيث أن استكمال الأداء لا يتحقق إلا بإتقان الأداء على الجزئيات. من ناحية ثانية، يشير المحك إلى مستوى أداء (standard) أو درجة قطع أو عبور (cut-score)، وهي المستوى أو الحد الأدنى الذي لو حققه الفرد يكون قد أتقن ذلك المحتوى أو تلك المهارة، كي يتمكن من القيام بالأداء اللاحق.

المعيار Norm

توزيع لدرجات نتجت عن أداء مجموعة معيارية على أداة قياس معينة. وبشكل أدق يعرف المعيار بأنه النقطة الوسيطة (الوسيط) لمجموعة من درجات مجموعة من الطلبة، والتي يقع 50 % من الدرجات فوقها، و 50 % منها تحتها.

العلامات الحدية (الهادية) Benchmark

الوصف التفصيلي للحد الأدنى المتوقع لأداء الطلبة في أعمار، أو مراحل دراسية أو مستويات نمائية معينة. يتم تطوير هذه العلامات بناءا على تحليل عينات من أعمال الطلبة وأدائهم. ويمكن استخدام تلك العلامات كنقاط مراقبة لضبط تقدم الطلبة نحو تحقيق الأهداف حسب المرحلة. وعادة ما يجري تحديد علامات حدية للقدرات المتوقعة في الرياضيات مثلا في الصفوف: 3، 7، 10 .

التقييم أو التقدير Assessment

عملية تجميع ووصف وتكميم المعلومات والبيانات المتعلقة بالأداء بقصد المساعدة في اتخاذ قرار ما. ويعرف التقييم بأنه أسلوب متعدد السمات ومتعدد الطرق Multitrait-multimethod وليس قياسا أحادي البعد One-dimentional، بمعنى أنه يتم فيه التركيز على أكثر من متغير وباستخدام أكثر من أسلوب (الاختبارات، المقابلات، والملاحظات، والإستبانات، وملفات الانجاز، ومقاييس التقدير وغير ذلك من أدوات القياس). والتقييم أكثر شمولا من القياس ويسبق التقويم وهو أقل شمولا منه. من ناحية أخرى، يقترن مفهوم التقييم عادة بالمجالات الإكلينيكية. حيث يميل الإكلينيكيون إلى استخدام هذا المفهوم في فحوصهم وتشخيصاتهم لاضطرابات الشخصية التي يعاني منها الفرد.

التقييم متعدد الأبعاد Multidimensional Assessment

التقييم الذي يتم فيه جمع بيانات عن مدى واسع من القدرات والمهارات كما جاء في نظرية هوارد غاردنر (Gardner, 1983 ; Nitko, 2001) حول الذكاءات المتعددة multiple intelligences. إذ تنص هذه النظرية على أن الشخص يتمتع بثمانية أنواع من الذكاء وليس واحدا. ويتفاوت الشخص الواحد في امتلاكه لأنواع الذكاء من حيث الدرجة. ومن أمثلتها الذكاء اللغوي، والمنطقي والرياضي، والمكاني، والجسمي، والموسيقي، والذكاء الشخصي، وذكاء التعامل بين الأشخاص وغيرها.

نظام التقييم Assessment System

مزيج من أنواع التقييم متعدد الأبعاد تستخدم معا بقصد الخروج بتقرير يتضمن معلومات شاملة، وموثوقة، وثابتة يتم استنادا عليها صنع

قرارات عن الطلبة والمدارس والمناطق التعليمية. ويتكون نظام التقييم من كل من: (1) التقييم معياري المرجع، (2) التقييم محكي المرجع و (3) التقييم الصفي.

التقييم محكي المرجع Criterion-Referenced Assessment

التقييم الذي يتم من خلاله مقارنة مستوى أداء الطالب بهدف تعلمي أو بأداء معياري محدد. وليس مقارنته بمستوى أداء الأفراد الآخرين. أي أن هذا النوع من التقييم يساعد في تحديد مستوى أداء الطلبة على أهداف ومستويات أكثر من مقارنة أدائهم مع أداء مجموعة معيارية محلية أو وطنية. وقد يصل جميع الطلبة أو ربما لا يصل أحد منهم مستوى الأداء المعياري.

من ناحية أخرى، تختلف تعريفات التقييم محكي المرجع بناءا على اختلاف طبيعة المجال السلوكي الذي تنسب إليه درجة الفرد، ومحتوى المجال، ودرجة تعقده. لذلك، يبرز هنا ثلاث أنواع من الاختبارات ذات العلاقة وهي: اختبارات مرجعية الهدف (Objective-Referenced)، واختبارات مرجعية المجال (Domain-Referenced)، واختبارات التمكن أو الإتقان (Mastery).

الاختبارات مرجعية الهدف Objective-Referenced Tests (ORT)

اختبارات تبنى على أساس مجموعة من الأهداف التعليمية المصوغة صياغة سلوكية، وتكون هناك مطابقة بين فقرات الاختبار والأهداف. وبناءا على الأداء على هذا النوع من الاختبارات، يصار إلى تصنيف المفحوصين إلى فئات. واحدة حققت الأهداف بدرجة عالية، وأخرى حققتها بنسبة معينة، وفئة لم تحقق. ولا يتطلب هذا النوع من الاختبارات تحديد مستوى الإتقان (المحك).

الاختبارات مرجعية المجال Domain-Referenced-Tests (DRT)

اختبارات تبنى في ضوء تحديد مجال سلوكي من المهارات والمتطلبات تحديدا دقيقا، ويتم انتقاء الفقرات التي يتكون منها المجال عشوائيا (عشوائية طبقية). وتكون درجة المفحوص على هذا الاختبار مقدمة لتقدير احتمال إجابة الفرد على مفردات المجال الشامل في وقت معين.

اختبارات التمكن أو الإتقان Mastery Test

اختبارات تبنى من أجل تحديد ما إذا كان الفرد قد اكتسب سلوكا يهدف البرنامج التعليمي أو التدريبي إلى تنميته لديه. ويساعد اختبار التمكن في اتخاذ قرارات تتعلق بإتقان الفرد لهدف تعليمي أو مهارة معينة أو مجال من المهارات. وجدير بالذكر أن اختبار التمكن يمكن أن يكون مرجعي الهدف أو مرجعي المجال. ويتطلب اختبار الإتقان تحديد المحك بالقول مثلا يجيب هذا الفرد 95 % من فقرات الاختبار إجابة صحيحة.

التقييم معياري المرجع Norm-Referenced Assessment

التقييم الذي يتم من خلاله مقارنة مستوى أداء الطالب مع أداء مجموعة أكبر تدعى (مجموعة معيارية)، قد تكون عينة وطنية تمثل تقاطعا عريضا ومتنوعا من الأفراد. ويمكن مقارنة الطلبة والمدارس والمناطق التعليمية مع المجموعة المعيارية. والهدف الأساس من هذا النوع من التقييم هو ترتيب الطلبة وليس قياس تحصيلهم تجاه بعض معايير الأداء.

التقييم الصفي Classroom Assessment

التقييم الذي يتم تطويره وتنفيذه وتقديره من قبل المعلم أو مجموعة من المعلمين بهدف تقويم أداء الطالب أو مجمل أداء طلبة صف في موضوع معين. وقد يكون التقييم الصفي معياري المرجع أو محكي المرجع. وبشكل

عام، يستخدم التقييم الصفي في تطوير التدريس وتمكين الطلبة من تحقيق مستويات عالية مرسومة.

التقييم على أساس الأداء Performance-Based Assessment

الملاحظة المنظمة والمباشرة والمستمرة على مدى فترة زمنية معينة لتقدير أداء المتعلم في تحقيق هدف تعلمي يتضمن خلق نتاجات Products معينة. ويكون هذا التقييم نشاطا تفاعليا بين المعلم والطالب، كجزء من عملية التعلم. كما أن المفروض أن يكون هذا التقييم أداء حقيقيا ومتصلا بالطالب والبيئة التعلمية. ويتم القيام بالتقييم باستخدام دليل تحليلي للتصحيح Rubricks لتعظيم موضوعيته. إنه اختبار القدرة على استخدام المعرفة في الأوضاع الحياتية.

عرف ستيجنز (Stiggins, 1992) التقييم المعتمد على الأداء بأنه عملية استخدام معايير الأداء لتحديد درجة تحقيق الطالب لهدف تحصيلي معين آخذين بعين الاعتبار عدد من العوامل والظروف المؤثرة بعملية التقييم كالتحيز الثقافي والعقبات اللغوية، وبيئة الامتحانات، والتحيز القائم على الامتحان وعينة المستهدفين والظروف المتعلقة بالبيانات وطرق جمعها.

محكات الأداء Performance Criteria

هي المستويات التي يتم تقييم أداء الطالب في ضوئها. وتساعد هذا المعايير المقيمين على تحقيق الموضوعية من ناحية، وتزويد المتعلم بمعلومات مهمة عن التوقعات التي تقود إلى الهدف أو الغاية المنشود تحقيقها من ناحية أخرى.

ملف الانجاز Portfolio

هو تجميع منظم ومنتظم لأعمال الطالب التي تم إنجازها أو عرضها، لتكون دليلا مباشرا على جهوده وتحصيله وتقدمه على مدى فترة

زمنية معينة طويلة نسبيا (عام دراسي). والمفروض أن يتم إشراك الطالب في اختيار محتويات الملف، كما يجب أن يتضمن ذلك معايير أداء موضوعية كمفتاح التصحيح التفصيلي rubrics من أجل تقييم مجمل أعمال الطالب. ويضم ملف الانجاز كافة أنواع الأعمال التي أنجزها الطالب بما في ذلك أشرطة الفيديو، وأقراص الحاسوب المرنة والمضغوطة، والسجلات والأعمال الكتابية والنشاطات بأنواعها.

التقييم بالبورتفوليو (التقييم بملفات الانجاز) Portfolio Assessment

يمكن التقييم بأسلوب ملفات الانجاز بعدة طرق. إذ يمكن تقييم كل جزء من أجزاء الملف أو تقييم كافة الأجزاء دفعة واحدة. وعادة ما يتم تقييم الملفات في ضوء دليل تصحيح rubric يتم تطويره لهذا الغرض. غالبا ما يتم استخراج المعايير من قبل الطلبة والمراجعين المشتركين في عملية تقويم التقدم في تحقيق الأهداف.

يناسب هذا النوع من التقييم الطلبة المحرومين ثقافيا أو الذين لديهم إعاقات لغوية. وتتيح البورتفوليو للمدرسين والموجهين إمكانية تقويم البرامج التعليمية، كما يتيح فرصة جيدة للطلبة أن يصبحوا شركاء مع مدرسيهم في عملية التقييم. كما يعطي أولياء الأمور فرصة التواصل والتفاهم والتعرف بفاعلية أكبر على أعمال التلاميذ. وتبرز ميزة التقييم بملفات الانجاز في أنه يغطي فترة زمنية أكبر من التي يغطيها الامتحان.

التقييم بالخريطة (بالصفحة) Profile Assessment

هو التقييم الذي يعتمد على تحليل تخطيط بياني وتراكمي لمستوى أداء الفرد (وأحيانا مجموعة من الأفراد) على سلسلة من أساليب التقييم، وخاصة الاختبارات التي تتضمنها بطارية الاختبارات.

التقييم بالنتاج Product

هو التقييم الذي يعتمد على تثمين النتاج المادي الملموس المتعلق بأداء أو مهمة معينة. من ناحية ثانية، يفترض أساسا أن النتاج الذي يتم تقييمه يعبر بالضرورة عن تعلم معين.

الروبريكس (قواعد التقدير) Rubrics

دليل تصحيح يستخدم في تقييم أعمال الطلبة (النتاج والأداء) كالاختبارات الصياغية أو غير الانتقائية. ويتضمن ذلك الدليل سلم أو عدد من المستويات تتألف عادة من أربعة (1– 4) أو ستة (1– 6)، مع وصف تفصيلي لكل مستوى.

تقييم الذات Self Assessment

هي عملية التقييم التي يتم إشراك الطالب فيها للقيام بمراجعة منظمة وتراكمية لأداءه، بغرض تطويره مستقبلا وعلى نحو أفضل. وتتم تلك المراجعة في ضوء محكات معيارية معرفة. كما يمكن أن يتضمن تقييم الذات مراجعة الفرد الناقدة لأعماله.

تقييم الأتراب (الرفاق أو الزملاء) Peer Assessment

تقييم غير رسمي يمكن أن يحدث في كل غرفة صفية، وفيه يقوم التلميذ بالنظر إلى عمل زميله أو أعمال زملائه في الصف. ويتم تحديد ما هو مناسب أو مقبول خلال جلسة نقدية محددة. ويبدو أن هذا التقييم يكون بمثابة تغذية راجعة لما يمكن أن يعمله الطالب أو يحسنه على نحو ما.

المواءمة Alignment

عملية ربط المحتوى ومعايير الأداء والتقييم والتعليم والتعلم في الصف الدراسي. وثمة استراتيجية مواءمة تتمثل في التطوير التدريجي

(خطوة – خطوة) لكل من: (1) معايير المحتوى، (2) معايير الأداء، (3) التقييم، (4) التدريس الصفي. ويتم عرض هذه النقاط بترتيب متتابع كما يلي: معايير المحتوى – معايير الأداء – التقييم – التدريس بغرض التعلم. وقد يحصل تداخل بين تلك العناصر. ولكن يبقى السؤال ما إذا كان التعليم والتعلم يدعم المعايير والتقييم؟

السجلات (دفاتر اليومية) Journals

هي السجلات التي تتعلق بشخصية الطالب وردود فعله تجاه جوانب تعلمية وممارسات تطويرية معينة. وتهدف هذه السجلات لتعظيم فرص التعلم.

التقييم التقليدي Traditional Assessment

التقييم الذي يتم فيه جمع بيانات ومعلومات عن المتعلم من خلال تحليل استجاباته على أدوات تقيس تعلمه في مجالات معرفية معينة. وتكون هذه الأدوات غالبا إنتقائية (من نوع الاختيار من متعدد، أو الصواب والخطأ، أو المزاوجة).

التقييم البديلي (المعتمد على البدائل) Alternative Assessment

التقييم الذي يطلب فيه من الطالب أن يقوم بإنتاج أو صياغة استجابة معينة تتعلق بسؤال أو مهمة. وتكون الاستجابة على هيئة شرح، أو ملفات انجاز، أو كتابة نصوص، أو تحرير مقالات وغيرها.

التقييم الأصيل (المرتبط بالواقع) Authentic Assessment

هو تقييم للأداء المتعلق بتناول مشكلة ذات صلة بواقع الحياة ومن خارج الإطار المدرسي، يتم اقتراحها من قبل مجموعة من الخبراء. ويتضمن ذلك تقييم قدرة الطالب على إيجاد حلول لتلك المسائل إلى مستوى

التمكن (الإتقان). كما تتطلب القدرة على هذا النوع من التقييم مدى واسعا من المعارف والمهارات ذات العلاقة بتلك المسائل.

ويتضمن التقييم الأصيل أنشطة منوعة مثل المقابلات الشفوية، ومهام حل المشكلة جماعيا، وإعداد بورتفوليو للكتابة وغيرها. ومن فوائد هذا التقييم تغيير دور التلاميذ ليصبح أكثر مشاركة ونشاطا في الأعمال الصفية من ناحية، كذلك تغيير دور كل من المدرسين وأولياء الأمور من ناحية أخرى.

مستوى التمكن (الإتقان) Mastery Level

هو المستوى الذي يصله الطالب في أدائه التعلمي. ويرتفع الطالب إليه تدريجيا من جزء من المطلوب تعلمه إلى معظمه حتى يصل إلى كل المطلوب.

الألفة بالتقييم Assessment Literacy

مدى امتلاك المعرفة المتصلة بالمبادىء الأساسية عن ممارسة التقييم الصوتي sound assessment كما في المصطلحات، وتطوير واستخدام منهجيات التقييم وتقنياته، وكذلك الإلمام بمعايير النوعية في التقييم ومعرفة البدائل للقياس التقليدي للتعلم.

المساءلة Accountability

أسلوب تقييمي يقوم به أفراد من المجتمع والمسؤولين ودافعي الضرائب. يتم بالاستفسار من القائمين على المدرسة عن جدوى إنفاق الموازنة المخصصة للمدرسة ومدى تحقق الأهداف وتقييم دور المديرين في مجال التعيينات والترقيات والمكافآت وغيرها. وتعد المساءلة أسلوب تقيمي يساعد في عمليات التطوير التربوي عموما وفي تحديد حاجات المدرسة. الأمر الذي يعمل على تسهيل عملية استغلال المصادر المتاحة

بعدالة وفاعلية أكثر. من ناحية أخرى، غالبا ما تجري المساءلة بشكل علني، وتتعلق في بحث مؤشرات متنوعة ومتعددة مثل مؤشرات كفاية العاملين، وحجم الصف الدراسي، والعدالة، وطرق التدريس والمنهاج، والتسرب، ومشاركة أولياء أمور الطلبة، وكذلك تحليل الدرجات، ومن ثم توظيفها لأغراض التطوير.

تحليل الفقرات Item Analysis

عملية فحص استجابات المفحوصين على فقرات الاختبار للحكم على مستوى نوعية كل فقرة. وأكثر المؤشرات التي يرتبط تحليل الفقرات في بحثها هي معامل صعوبة الفقرة، ومعامل تمييزها، وكذلك معامل فعالية أو جاذبية مموهات تلك الفقرة.

معامل صعوبة الفقرة Item Difficulty Index

هي نسبة المفحوصين الذين أجابوا تلك الفقرة إجابة صحيحة. وتمتد قيم معامل الصعوبة بين الصفر والواحد (لا يمكن أن تكون مقدارا سالبا). وتزداد درجة صعوبة الفقرة كلما يقل معامل صعوبتها. وفي الاختبارات معيارية المرجع، تفضل الفقرات التي تمتد معاملات صعوبتها بين (40.) و (60.) .

معامل تمييز الفقرة Item Discrimination Index

قدرة الفقرة على التمييز بين المفحوصين من فئة ذوي الأداء المنخفض وفئة ذوي الأداء المرتفع في إجابتهم على الفقرة. وتمتد قيم معامل التمييز بين (-1.0) و (+1.0) . ومع ضبط العوامل الأخرى، فإنه كلما زادت قيمة معامل تمييز الفقرة الموجب كانت الفقرة أفضل. أما الفقرة ذات التمييز السالب، فينصح بتعديلها وإعادة كتابتها.

المموهات Distracters

هي البدائل الخاطئة التي تحويها الفقرات الانتقائية من نوع فقرات الاختيار من متعدد أو فقرات المزاوجة. والمفروض أن يتم اشتقاق مموهات الفقرة بناءا على الإجابات التي يمكن أن يظهرها الطلبة ضعيفي التحصيل. وعادة ما يتم صيغتها من خلال إجابات مجموعة من الطلبة على عدد من الأسئلة الصياغية.

جاذبية المموه Distracter Attractiveness

قدرة البديل الخاطىء (المموه) في الفقرة الانتقائية على جذب المستجيبين من فئة الأداء المنخفض على اختياره. وتتراوح قيمة جاذبية المموه بين (-1.0) و (+1.0). وكلما كانت الجاذبية سالبة وكبيرة كان المموه أكثر جاذبية وفعالية وينصح بالإبقاء عليه في الفقرة. أما المموه الذي تكون جاذبيته موجبة أو صفرا، فيجب تعديله وتطويره، أو حتى استبداله، ليكون أقدر على جذب نسبة أكبر من المستجيبين من فئة ذوي الأداء المنخفض.

صدق الفقرة Item Validity

هو حاصل ضرب الانحراف المعياري لدرجات الممتحنين على تلك الفقرة بمعامل ارتباط الدرجات على تلك الفقرة بالدرجات الكلية للمتحنين على المحك الخارجي.

ثبات الفقرة Item Reliability

هو حاصل ضرب الانحراف المعياري لدرجات الممتحنين على تلك الفقرة بمعامل ارتباط الدرجات على تلك الفقرة بالدرجات الكلية على المحك الداخلي (درجات المفحوصين الكلية على الاختبار).

حساسية الفقرة Item Sensitivity

قدرة الفقرة في الاختبار محكي المرجع على التمييز بين المفحوصين الذين تلقوا تدريبا أو تعليما أو إرشادا معينا وهؤلاء الذين لم يتلقوا. وتمتد قيمها بين (−1) و (+1).

درجة التمكن Mastery Score

أقل درجة عبور (نجاح أو قبول) أو درجة قطع Cut-Score أو Cutoff على اختبار محكي المرجع. ويصنف الأفراد الذين يحصلون على هذه الدرجة أو يتجاوزونها بأنهم متمكنون من مادة دراسية معينة أو مهارة محددة. وهناك عدة طرق تستخدم في تحديد تلك الدرجة، يعتمد بعضها على بحوث وتجارب، و تؤول كل طريقة منها إلى نتيجة مختلفة عن الأخرى.

المتغير variable

صفة أو خاصية من خواص شيء أو فرد، قد تأخذ أكثر من قيمة أو مستوى في الظروف والأوقات والحالات المختلفة. ومن الأمثلة على ذلك متغيرات: العمر، والجنس، واللون، والوزن، والطول، وغيرها. فالجنس يكون ذكراً أو أنثى، ويأخذ العمر قيماً متعددة، فقد يكون سنة، أو عشر، أو أكثر أو أقل من ذلك. وتتنوع المتغيرات ومستوياتها بحسب طبيعة البحث وأهدافه، ومن أمثلة المتغيرات في المجالات التربوية والنفسية والاجتماعية التحصيل، والذكاء، وطرق التدريس، ومستوى القلق، والاختيار الدراسي، ودرجة الرضا، وفعالية الأداء الإداري. من ناحية أخرى، عندما يتشابه مجموعة من الأفراد في خاصية معينة، تكون هذه الخاصية المجموعة **ثابتا (Constant)** بالنسبة لتلك المجموعة، أي أنها ليست متغيراً.

المتغير المستقل Independent variable

ويعرف بالمؤثر أو المسبب، وهو المتغير الذي يتم التحكم به من قبل الباحث في التجارب الطبيعية أو النفسية أو الاجتماعية بهدف التعرف على تأثيراته في نتيجة التجربة. ويدعى المتغير المستقل أحياناً **بالمتغير التجريبي**.

المتغير التابع Dependent variable

ويعرف بالنتيجة، أو المتغير الذي يقع عليه التأثير من المتغير المستقل. فمثلاً عند دراسة أثر اختلاف طريقة التدريس على تحصيل طلبة الصف السادس في الرياضيات، فإن طريقة التدريس هي المتغير المستقل، بينما يكون التحصيل متغيراً تابعاً. ويعرف المتغير التابع أحيانا **بالمحك** **Criterion**.

البيانات(المتغيرات) الكمية Quantitative Data (Variables)

هي نوع من البيانات التي يمكن التعبير عنها كمياً، أو يمكن تمييزها بالدرجة أو التكرار أو الكمية، كالتحصيل الذي يعبر عنه بدرجات الاختبارات (90 ، 88 ، 74، 56) أو الغياب الذي يعبر عنه بالتكرارات (6 مرات، 3 مرات، 11 مرة) خلال فصل دراسي معين، وأطوال التلاميذ في الصف الثالث الأساسي (110 سم، 115 سم، 98 سم). ويمكن أن تكون قيم المتغيرات الكمية مرتبة في أولويات من مثل: الأول، أو الثاني، أو العاشر، أو العشرون.

البيانات (المتغيرات) النوعية Qualitative Data (Variables)

هي نوع من البيانات التي تصنّف الأفراد أو الأشياء حسب النوع، ولهذا فهي فاقدة صفة الترتيب والأفضلية، ومثال ذلك متغير الجنس الذي يصنّف الأفراد إلى ذكور أو إناث، ومتغير الفرع الدراسي الذي يصنف

الطلبة إلى أدبي أو علمي أو مهني، وهكذا. وتستخدم الأرقـام فـي هـذه المتغيرات لغايات التصنيف فقط، ولا تعكس مدلولات كمية. فعندما للذكور بالرقم (1) وللإناث بالرقم (2)، فإن العددين (1) و(2) يستخدما للترميـز فقط وليس لأي منهما معنى تفضيلي لأحدهما على الآخر.

البيانات (المتغيرات) المستمرة أو المتصلة Continuous Data

هي البيانات التي تأخذ عناصرها أي قيمة (الكسور مثلا) في متصل معين. وهي قابلة للعد والترتيب والقياس كما في الوزن، والطول، والدخل، وكمية السكر في لتر من عصير البرتقال، وكمية الزمن اللازم لرياضي متمرس لقطع مسافة مقدارها كيلومترين.

البيانات (المتغيرات) المتقطعة أو الوثابة Discrete Data

هي البيانات التي تأخذ عناصـرها قيمـا صـحيحة محـددة (غيـر كسرية).وهذه القيم بمثابة رموز ومسميات، كالجنس الذي يكون ذكـراً أو أنثى فقط، ولا توجد بينهما قيمة أخرى، وعدد تلاميذ الصف (20 طالبا) أو عدد أفراد الأسرة (ستة أفراد) وما شابه. ففي هذه الحـالات لا يعقـل أن يكون عدد الطلبة يحوي كسرا.

1- مستوى القياس الاسمي Nominal scale of Measurement

هو المستوى الأدنى للقياس، ويستخدم مع المتغيرات النوعية، حيـث يتولى القياس الاسمي تصنيف الأفراد أو الأشياء في عدة مجموعات وفقـاً لبعض الخصائص النوعية كتوزيعهم حسب جنسهم (ذكـور وإنـاث)، أو حسب مناطقهم السكنية (جنوب، شرق، شمال، غـرب)، أو حـسب لـون العيون (أزرق، أخضر، عسلي، أسود)، أو حسب الرقم الوطني، أو حسب لون البشرة، أو في ضوء الأرقام المكتوبة على صدور اللاعبين، والـرقم الجامعي.

2-مستوى القياس الرتبي **Ordinal Scale of Measurement**

يُعتبر هذا النوع من المقاييس تالياً من حيث المستوى للمقاييس الإسمية فهو أعلى منها. فبالإضافة إلى تصنيف الأفراد والأشياء في مجموعات متمايزة، يرتب الأفراد والأشياء تصاعدياً أو تنازلياً في صفة أو خاصية معينة، وعندما تعطى الأرقام للأشياء والأفراد وفقاً لهذا المقياس، فإنها تمثل كميات معينة. كما أن المسافات الفاصلة بين رقم وآخر لا يشترط أن تكون متساوية.

ومن الأمثلة على المقياس الرتبي، درجات المدرسين حسب الكفاءة التدريسية، ورتب التلاميذ حسب نشاطهم في الصف، وترتيب الفرق الرياضية في لعبة معينة حسب مستوى الجدارة والكفاءة في تلك اللعبة، وكذلك رتب مجموعة من الضباط.

3-مستوى القياس الفئوي **Interval scale of Measurement**

يُعتبر هذا النوع من المقاييس أعلى مستوى من المقياسين الاسمي والرتبي، ويمتلك خاصية الفواصل أو المسافات المتساوية التي تفصل بين كل درجتين متجاورتين. هذا يشير إلى إمكانية إجراء عمليات حسابية لإيجاد الفروق، فعندما تكون درجات أربع طلاب هي 55، 60، 70، 75، فإنه يمكن القول بأن الفرق بين (55) و (60) هو (5) ويساوي الفرق بين (75) و (70). كما يجوز القول بأن الفرق بين العلامتين (75) و (60) يساوي ثلاثة أمثال الفرق بين العلامتين (60) و (65). و لكن لا يمكن القول بأن العلامة (70) تساوي ضعف العلامة (35). كما أن الصفر في هذا المستوى افتراضي وليس حقيقياً، بمعنى أن الصفر قيمة كباقي القيم.

مستوى القياس النسبي **Ratio Scale of Measurement**

هو أعلى مستويات القياس، ويتميز بخصائص جميع المقاييس السابقة، إضافة إلى وجود الصفر المطلق فيه (بمعنى أن الصفر المطلق هو صفر

حقيقي ويعني غياب الصفة). فعندما نقول إن فلاناً يملك صفراً من النقـــود فهذا يعني أنه لا يملك نقوداً أبداً. ويصح في هذا المستوى تطبيـــق مفهـــوم النسبة، كما يمكن إجراء كافة العمليات الحسابية.

القيمة المتطرفة Outliers

هي قيم أو قيم يحويها توزيع من البيانات الكمية وتختلف بدرجة غير عادية عن مجمل البيانات (صغرا أو كبرا). أو هي القيم التي تختلف عن المتوسط الحسابي للتوزيع بمقدار انحرافين معياريين أو أكثر. وتظهر تلك القيم في التوزيع نتيجة خطأ في عملية القياس، يفضل الانتباه جيدا لمدى تأثيرها على بعض مؤشرات النزعة المركزية والتشتت وغيرها. من ناحية أخرى، يخشى أن تكون تلك القيم حقيقية في التوزيع، أي أنها تنتج عن قياس حقيقي لسلوك أو أداء معين، لذلك فإنه لا ينصح بالتسرع بحذفها.

المتوسط الحسابي Arithmetic Mean

نقطة في توزيع من القيم الكمية، يعرف أحيانا بالمعدل Average. من ناحية أخرى، يساوي مجموع انحرافات كافة القيم عن المتوسط الحسابي صفرا. ويتم حسابه كميا بإيجاد حاصل قسمة مجموع القيم على عددها، لذلك فإن قيمته تعتمد على كافة القيم في التوزيع، وبذلك لا ينصح باستخدام المتوسط الحسابي كمؤشر نزعة مركزية لتوزيع ملتو للبيانات أو في حالة وجود قيم متطرفة.

الوسيط Median

نقطة في توزيع من القيم الكمية. وهي النقطة التي يكون (50 %) من القيم المرتبة تصاعديا أو تنازليا قبلها و (50 %) بعدها. ويعرف الوسيط أيضا بأنه المئين (50). ولا يتأثر مقداره بالقيم المتطرفة في التوزيع، لهذا

فهو مقياس النزعة المركزية المفضل استخدامه في حالة وجود قيم متطرفة أو في حالة التوزيعات الملتوية Skewed Distributions.

المنوال Mode

القيمة ذات التكرار الأكبر في توزيع من البيانات. وقد يحوي بعض التوزيعات أكثر من منوال. عند ذلك يدعى ذلك التوزيع بثنائي أو ثلاثي أو حتى متعدد المنوالات. وقد يكون المنوال مقدارا كميا أو صفة نوعية. فنقول أن أكثر الدرجات تكرارا هي 75 (كميا). أو أكثر الأسماء تكرارا في مجتمع ما هي محمد (نوعيا).

التماثل Symmetry

حالة تماثل توزيع مجموعة من البيانات حول محور معين. وغالبا ما ينسب التماثل إلى توزيع قيم العينة على جانبي المحور العمودي. والتوزيع الطبيعي هو من التوزيعات المشهورة التي تتميز بصفة التماثل.

الالتواء Skewness

هي حالة عدم التماثل Asymmetry لتوزيع عينة أو مجتمع من البيانات الكمية. فقد تتجمع بيانات في طرف التوزيع أكثر من تجمعها في الطرف الآخر. وهناك نوعان من الالتواء: التواء نحو اليمين وآخر نحو اليسار. ويمكن تحويل البيانات الكمية الملتوية إلى بيانات تتميز بالتماثل، أو أنها تتوزع توزيعا طبيعيا بدرجة معقولة، وذلك بإخضاعها إلى عمليات تحويل Transformation مناسبة.

التوزيع الملتوي التواءا موجبا Positively Skewed Distribution

توزيع غير اعتدالي (غير متماثل asymmetric) لمجموعة من القيم (الدرجات)، فيه عدد كبير من الدرجات المنخفضة، وعدد أقل نسبيا

من العالية. ويكون المتوسط الحسابي لهذه الدرجات أكبر من الوسيط وهذا أكبر من المنوال. من ناحية أخرى، يعتبر الاختبار التحصيلي الذي تتوزع درجات المفحوصين عليه بأنه اختبار صعب نسبيا. ويعرف هذا النوع من التوزيع باسم **التوزيع الملتوي نحو اليمين.**

التوزيع الملتوي التواءا سالبا Negatively Skewed Distribution

توزيع غير اعتدالي (غير متماثل asymmetric) لمجموعة من القيم (الدرجات)، فيه عدد كبير من الدرجات العالية، وعدد أقل نسبيا من المنخفضة. ويكون المتوسط الحسابي لهذه الدرجات أصغر من الوسيط وهذا أصغر من المنوال. من ناحية أخرى، يعتبر الاختبار التحصيلي الذي تتوزع درجات المفحوصين عليه بأنه اختبار سهل نسبيا. ويعرف هذا النوع من التوزيع باسم **التوزيع الملتوي نحو اليسار.**

تحويل البيانات Data Transformation

المقصود هنا تحويل البيانات الملتوية إلى التوزيع الطبيعي. فعندما تكون البيانات ملتوية إيجابيا (ملتوية نحو اليمين)، يمكن تحويلها إلى بيانات طبيعية بإيجاد مقلوب، أو اللوغاريتم أو إيجاد الجذر التربيعي لكل قيمة. أما في حالة البيانات الملتوية سلبا (ملتوية نحو اليسار)، فيمكن تحويلها إلى بيانات طبيعية بإيجاد مربع أو مكعب كل قيمة منها. من ناحية أخرى، هناك نوعان من تحويل البيانات (خطي وغير خطي) ويظهر شرح لكل منهما في موقع لاحق من هذا الفصل.

التفلطح Kurtosis

هي مقياس تدبب توزيع البيانات الكمية. فإذا كان التوزيع مدببا، يكون تفلطح التوزيع موجبا، بينما إذا كان رأس التوزيع مسطحا فإن

تفلطحه سالبا. وهناك معادلة رياضية لقياس درجة تفلطح التوزيع البياني لبيانات كمية.

التوزيع الطبيعي Normal Distribution

توزيع متماثل لقيم حول متوسطها الحسابي، ويأخذ شكل الجرس. وعادة ما تتركز تلك القيم قرب متوسطها الحسابي وتقل تكراراتها كلما ابتعدنا عن المتوسط في الاتجاهين طبقا لصيغة رياضية محددة. من ناحية أخرى، يقع 68 % من القيم بين نقطتين يبعد كل منهما انحرافا معياريا واحدا على جانبي المتوسط. كما يقع 95 % و 99 %من القيم بين نقطتين يبعد كل منهما انحرافين معياريين أو ثلاث انحرافات معيارية على الترتيب على جانبي المتوسط. من ناحية أخرى، فإن مقدار كل من الالتواء والتفلطح للتوزيع الطبيعي صفرا.

التشتت Variation (Dispersion)

حالة تباعد البيانات عن بعضها البعض، أو تباعدها عن أحد مؤشرات النزعة المركزية (المتوسط الحسابي مثلا). فعندما يكون التشتت كبيرا، فإن البيانات تتبعثر عن بعضها بشكل واسع. وعندما يكون التشتت قليلا، تبدو البيانات متقاربة أو متجمعة Clustered . ولقياس مقدار تشتت البيانات هناك عدد من المؤشرات مثل المدى و الانحراف المعياري والتباين ونصف المدى الربعي وغيرها.

المدى Range

أحد مقاييس تشتت البيانات الكمية. وهو الفرق بين أكبر قيمة وأصغر قيمة في التوزيع (لذلك فهو كمية موجبة دائما). يمتاز المدى بسهولة حسابه، هذا على الرغم من أنه يتعامل فقط مع قيمتين من قيم

التوزيع ويتجاهل باقي القيم. لذلك، فهو يتأثر بالقيم المتطرفة، ولا ينصح باستخدامه في حالة وجودها.

نصف المدى الربعي Semi-Inter-Quartile Range

هو نصف الفرق بين الربعي الأعلى (الثالث) والربعي الأدنى (الأول) لتوزيع البيانات الكمية. وبشكل عام، فإن قيمة نصف المدى الربعي أقل من قيمة المدى وأقل تأثرا بالقيم المتطرفة. الأمر الذي يمكن اعتباره مقياسا جيدا لتشتت التوزيعات الملتوية.

الربعيات Quartiles

هي القيم التي تقسم عينة من البيانات الكمية إلى أربعة أقسام متساوية، كل منها يحوي عددا متساويا (بدرجة كبيرة) من البيانات. لذلك يوجد ثلاثة ربعيات فقط لكل مجموعة من البيانات هي الربعي الأول (Q1)، والربعي الثاني (Q2)، والربعي الثالث (Q3).

المئينات Percentiles

هي القيم التي تقسم عينة من البيانات الكمية إلى مائة قسم متساوية، كل واحد يحوي عددا متساويا (بدرجة كبيرة) من البيانات. فمثلا يقع 30 % من البيانات تحت المئين 30 (P30). ويقع 50 % منها تحت المئين 50 (P50) أو ما يعرف بالوسيط.

الانحراف المعياري Standard Deviation

أحد مقاييس تشتت البيانات الكمية. وهو الجذر التربيعي لمجموع مربعات انحرافات القيم عن متوسطها الحسابي مقسوما على عدد القيم التي يتألف من مجتمع البيانات. أو مقسوما على عدد القيم التي تتألف من عينة البيانات مطروحا منها واحد.

التباين Variance

أحد مقاييس تشتت البيانات الكمية. وهو مجموع مربعات انحرافات القيم عن متوسطها مقسوما على عدد القيم التي يتألف منها مجتمع البيانات. أو مقسوما على عدد القيم التي تتألف من عينة البيانات مطروحا منها واحد. والتباين هو مربع الانحراف المعياري، وهو أكثر مقاييس التشتت استخداما في التحليلات الاحصائية واختبار الفرضيات.

معامل الارتباط Correlation Coefficient

مؤشر كمي لقوة واتجاه العلاقة بين مجموعتين من القيم الكمية، تتراوح قيمته بين (−1) و (+1). وتعرف العلاقة بالتامة عندما تصل قيمة معامل الارتباط واحد (سواء موجبة أو سالبة). وتشير القيمة صفر إلى انعدام العلاقة (خطية أو غير خطية) بين مجموعتي القيم. ويستخدم معامل الارتباط في تقدير معاملي صدق الاختبار وثباته. من ناحية أخرى يجري إيجاد مربع مقدار معامل الارتباط (معامل التحديد **Coefficient of Determination**) لتفسير مقدار الارتباط أو هو التباين المشترك بين متغيرين.

الدرجة الخام Raw Score

هي الدرجة التي يحصل عليها المفحوص على اختبار معين. كما أنها تعرف بمجموع الدرجات التي حصل عليها الطالب على أجزاء الاختبار نتيجة اتباع مفتاح تصحيح معين، لذلك الاختبار. وقد تحوي هذه الدرجة أخطاء تعزى إلى مصادر تتعلق بالاختبار، أو بالمفحوص، أو بتطبيق الاختبار وغيرها. وتعرف الدرجة الخام أحيانا بالدرجة الملاحظة Observed Score ، أو المحصلة Obtained Score. وتتألف الدرجة

الخام (x) من مكونين هما الدرجة الحقيقية (T) والدرجة الخطأ (E) حسب المعادلة:($X = T + E$).

الدرجة الحقيقية True Score

هي الدرجة التي لا تحوي جزءا خطأ؛ وعادة ما تعرف بأنها معدل الدرجات التي يحصل عليها المفحوص إذا عرض عليه الاختبار نفسه عددا لا نهائي من المرات (بافتراض عدم حصول أي تعلم). ومن ناحية إحصائية، يمكن تقدير الدرجة الحقيقية على هيئة فترة ثقة Confidence Interval، وبدلالة كل من قيمة الخطأ المعياري للقياس. من ناحية أخرى، فإن الدرجة الحقيقية مفهوم نظري، ومن النادر الحصول عليه عمليا.

ثبات الاختبار Test Reliability

المدى الذي يقيس الاختبار ما يهدف إليه على مدى فترات زمنية ومناسبات مختلفة. ويشير الثبات إلى مدى دقة المقياس، واستقراره، وخلوه من الأخطاء العشوائية. كما يعرف أيضا بأنه إمكانية توليد أو تحصيل Reproduceability مجموعة الدرجات جراء تطبيق الاختبار تحت ظروف وأوقات متنوعة. وبالرغم من أن الثبات شرط ضروري للصدق، إلا أنه ليس كافيا.

معامل الثبات Reliability Coefficient

معامل الارتباط المصمم لتقدير ثبات الاختبار من خلال: (أ) الارتباط بين الدرجات على الأشكال المتكافئة للاختبارات، أو (ب) الارتباط بين الدرجات بين نصفي الاختبار المتكافئين والمصححان لطول الاختبار، أو (ج) الارتباط بين الدرجات الناتجة من تطبيق الاختبار الواحد مرتين تفصلهما فترة زمنية معينة، أو (د) معامل الاتساق الداخلي للإختبار.

ويتوقع أن لا يقل معامل ثبات الاختبارات التحصيلية الصفية عن (0.60)، ويجب أن يفوق (0.80) في حالة مقاييس الاستعداد والاختبارات التحصيلية المقننة.

خطأ القياس Error of Measurement

مقدار الفرق بين الدرجة التي يحصل عليها المفحوص (الدرجة الخام أو الملاحظة) والدرجة المفترضة (الحقيقية). ويقع خطأ القياس في نوعين: عشوائي Random، ومنتظم Systematic.

الخطأ العشوائي Random Error

هو مجمل ما يؤثر على الدرجات الخام للمفحوصين من مصادر مختلفة يصعب ضبطها والحد من تأثيرها. ويؤثر هذا النوع من الخطأ على درجات المفحوصين بمقادير واتجاهات مختلفة. فإذا وجد صوت ضجيج خارج قاعة الامتحان، فإنه يتوقع أن يتأثر مستوى أداء كل مفحوص بدرجة مختلفة وبأسلوب يختلف من شخص إلى آخر. من ناحية ثانية، يؤثر الخطأ العشوائي على كل من صدق الاختبار وثباته.

الخطأ المنتظم Systematic Error

مقدار ما يؤثر على درجة المفحوص من مصادر يمكن معرفتها وبالتالي ضبطها. وخير مثال على هذا النوع من الخطأ، مقدار الزيادة الثابتة أو النقص الثابت التي يظهرها ميزان معين في كل مرة يستخدم فيها. ولا يؤثر هذا النوع على ثبات الاختبار، بل يلوث صدقه.

الخطأ المعياري للقياس Standard Error of Measurement

هو مقدار التذبذب fluctuation المتوقع للدرجة الخام حول الدرجة الحقيقية. فمثلا، لن تختلف الدرجة الملاحظة بأكثر من خطأ معياري واحد

أكثر أو أقل من الدرجة الحقيقية 68 % من الوقت. وبعبارة أخرى، فإن 95 % من الوقت لا تختلف الدرجة الملاحظة بمقدار أكثر من خطأين معياريين أقل أو أكثر من الدرجة الحقيقية للمفحوص. يعتمد مقدار الخطأ المعياري للقياس على مدى تشتت الدرجات (الانحراف المعياري) ومعامل ثباتها. فهو يزداد كلما قل مقدار معامل ثبات الاختبار.

الخطأ المعياري للتقدير Standard Error of Estimate

هو الانحراف المعياري المعتمد على الفروق بين الدرجات الملاحظة والدرجات المتنبأة من خلال معرفة معامل الارتباط بين الدرجات المتنبأة والدرجات على المحك. وكلما زاد مقدار معامل الارتباط، قل مقدار الخطأ المعياري في التقدير، وبالتالي زادت درجة دقة التقدير.

ثبات الاستقرار Stability Reliability

أو ما يعرف بطريقة الاختبار وإعادة الاختبار Test - Retest Method ويمكن تقدير ثبات الاستقرار من خلال تطبيق الاختبار موضع البحث على مجموعة من الأشخاص، ثم إعادة تطبيق الاختبار ذاته على المجموعة نفسها في وقت لاحق، ويتبع ذلك حساب معامل الارتباط بين درجات أفراد المجموعة على ذلك الاختبار في الفترتين.

ثبات التكافؤ Equivalence Reliability

معامل الثبات الذي يمكن تقديره من خلال إعطاء شكلين متكافئين بالمحتوى والمتوسطات والتباينات لاختبار معين لنفس المجموعة وبنفس الوقت (بعد فاصل زمني قصير)، وحساب معامل الارتباط بينهما.

ثبات الاستقرار والتكافؤ Stability-Equivalence Reliability

معامل الثبات الذي يمكن تقديره من خلال إعطاء شكل معين من أشكال الاختبار، وبعد فترة زمنية أطول نسبيا، يتم تطبيق الشكل الآخر للاختبار، ويتم بعد ذلك حساب معامل الارتباط بين مجموعتي الدرجات.

مقاييس الاتساق الداخلي Internal Consistency Measures

ثبات هي معاملات تستخدم في تقدير درجة تجانس فقرات الاختبار، أو أنها تعكس مدى ترابط استجابات المفحوصين على الفقرة الواحدة مع درجاتهم على الاختبار ككل.

التجزئة النصفية Split Half Reliability

أحد أساليب تقدير درجة ثبات الاتساق الداخلي لمقياس ما من خلال تطبيقه مرة واحدة على عينة ممثلة من المفحوصين. ويعتمد هذا التقدير على معامل الارتباط بين نصفي الاختبار (فردي ، زوجي مثلا). وهناك أسلوبان رياضيان يتعاملان مع هذا التقدير (1) أسلوب سبيرمان -براون ويفترض تساوي تباين الدرجات على نصفي الاختبار، و(2) أسلوب جوتمان عندما لا يتساوى تباين النصفين.

تقديرات كودر-ريتشادسون Kuder-Richardson Measures

أسلوب يستخدم في تقدير درجة ثبات الاتساق الداخلي لمقياس ما من خلال تطبيقه مرة واحدة على عينة ممثلة من المفحوصين. وتستخدم هذه التقديرات في حالة الاختبارات التي تتكون من فقرات تصحح ثنائيا (صفر ، واحد). وهناك تقديران للثبات لكودر- ريتشاردسون هما (KR-20) و (KR-21).

طريقة معامل كرونباخ ألفا Cronbach Alpha

أسلوب يستخدم في تقدير درجة ثبات الاتساق الداخلي لمقياس ما من خلال تطبيقه مرة واحدة على عينة ممثلة من المفحوصين. وتستخدم طريقة ألفا مع كافة الاختبارات. وتنطلق أهميتها من أنها تصلح لحساب ثبات المقاييس والاختبارات التحصيلية والشخصية والاتجاهات وغيرها.

ثبات الفرق بين درجتين Reliability of Difference Scores

ثبات الفرق بين درجتي مجموعة من المفحوصين على مقياسين أو أكثر (التوافق الشخصي والقلق مثلا). كما يستخدم هذا المؤشر في حالة المقارنة بين درجات متدربين قبل البرنامج وبعده.

ثبات التصحيح Scoring Reliability

هو مدى الارتباط (معامل الارتباط) بين مجموعتي درجات مجموعة من المفحوصين يقوم بتقدير كل منها (1) مصحح مرتين متتابعتين، أو (2) مجموعة من المصححين المستقلين. هذا بسبب احتمالية حدوث خطأ في تقدير الدرجات على الاختبارات المقالية، أو تحليل السلوك على الأدوات الإسقاطية ومقاييس التصنيف الذي ربما يرجع إلى الأشخاص الذين قاموا بالتصحيح أو التصنيف.

صدق الاختبار Test Validity

مدى ما يستطيع الاختبار أو المقياس أن يحقق من الأهداف التي بني من أجلها. ويتم التحقق من الصدق عمليا (صدق المحتوى)، أو منطقيا (صدق البناء) أو ارتباطيا (الصدق المرتبط بالمحك). ولا يعد الحكم على صدق الاختبار مسألة مطلقة، بل هي مسألة نسبية وموقفية. كما أنه لا يمكن

القول بأن اختبارا ما كامل الصدق أو أنه عديم الصدق، بل يمكن القول بأن هذا الاختبار عالي الصدق أو متوسط أو منخفض الصدق.

صدق المحتوى Content Validity

الدليل المنطقي Logical evidence على أن محتوى فقرة الاختبار مناسب للغرض الذي بني ذلك الاختبار من أجله. ويستخدم هذا النوع من الصدق أساسا مع اختبارات التحصيل. ويتم التحقق من درجة صدق المحتوى من خلال نتائج التحكيم الذي يقوم به مجموعة الخبراء اعتمادا على جدول المواصفات، ولا يتضمن التحقق منه استخدام أساليب إحصائية. من ناحية أخرى، يتصل بصدق المحتوى كل من الصدق الشكلي، والعيني، والمنهاجي، والصدق التدريسي.

الصدق الشكلي Face Validity

ويعرف أحيانا بالصدق الظاهري، وهو مظهر مصطنع لصدق الاختبار، ويعرف بمدى اعتقاد أو حكم المفحوصين على أن الاختبار مفيد ويقيس ما بني من أجله.

الصدق العيني Sampling Validity

درجة تمثيل الاختبار التحصيلي للمحتوى التعلمي موضع البحث. وهو أحد أشكال صدق المحتوى.

الصدق المنهاجي Curricular Validity

مدى اتساق فقرات الاختبار بأهداف منهاج مؤسسة تربوية كما صيغت بشكل رسمي. وهو أحد أشكال صدق المحتوى.

الصدق التدريسي Instructional Validity

مدى قياس فقرات الاختبار لمحتوى معين أو مهارة معينة كما تم تقديمه، أو تدريسه، أو عرضه. وهو أحد أشكال صدق المحتوى.

الصدق المرتبط بمحك Criterion-Related Validity

هو صدق المقياس المعتمد على معامل الارتباط بين الدرجات على المقياس والدرجات على المحك. ويعرف أحيانا بالصدق العملي Empirical Validity. ويقع هذا الصدق في نوعين من الصدق هما التزامني (التلازمي) والتنبؤ.

الصدق التزامني Concurrent Validity

الصدق المرتبط بالمحك عندما يتم حساب معامل الارتباط بين كل من الدرجات على المقياس والدرجات على المحك عندما يتم استخراجهما في نفس الوقت.

صدق التنبؤ Predictive Validity

الصدق المرتبط بالمحك والذي يتم تقديره من خلال معامل الارتباط بين درجات مجموعة من الأفراد على الاختبار ودرجاتهم على اختبار آخر (المحك) يتم تطبيقه مستقبلا.

تلوث المحك Criterion Contamination

تأثير معرفة مستوى أداء الفرد (المتقدم لدخول برنامج تدريبي معين مثلا) على الاختبار المتنبىء على الدرجة التي يستحقها على المحك. فإذا عرف مدرب البرنامج درجات الأفراد على اختبار القبول، فإنه ربما يتأثر تقديره لدرجات الأفراد على نتائج التدريب (المحك). وهذا التلوث ربما يعمل على إحداث زيادة مضللة على العلاقة بين المتنبىء والمحك عندما

يبدأ المدرب بالتعامل مع الأفراد الذين درجاتهم عالية نسبيا على المتنبىء على أنهم متفوقون حقا، ويبدي نحوهم اهتماما خاصا. وقد يعمل من جهة ثانية على تقليل مقدار العلاقة بين المتنبىء والمحك، إذا عرف المدرب بعضا ممن درجاتهم منخفضة، وبدأ التركيز عليهم وبذل جهد مميز من أجلهم.

صدق البناء Construct Validity

الدرجة التي يمكن بها أن يقيس الاختبار السمة الافتراضية (غير الملاحظة) وتفسر سلوكا. ويعتمد صدق البناء على كل من الأدلة العملية والمنطقية لمدى الارتباط بين الاختبار والنظرية ذات العلاقة. ويرتبط هذا النوع من الصدق بالسمات النفسية والقدرات العقلية.

الصدق التقاربي Convergent Validity

معامل الارتباط بين الدرجات على مقياس سمة معينة والدرجات على اختبارات أخرى تقيس سمات مشابهة. ويعتبر معامل الارتباط الموجب والعالي مؤشر على وجود الصدق التقاربي.

الصدق التمايزي Discriminant Validity

معامل الارتباط بين الدرجات على مقياس سمة معينة والدرجات على مقياس يقيس سمة مختلفة أو متعاكسة. ويعتبر معامل الارتباط السالب والعالي مؤشر على وجود الصدق التمايزي.

الصدق الذاتي Self Validity

الصدق الذاتي للاختبار هو الجذر التربيعي لمعامل الثبات. ولا يعتبر هذا الصدق بديلا لأي من مؤشرات الصدق المعروفة، ويعرف أحيانا بالصدق الهروبي.

التحليل العاملي Factor Analysis

أسلوب رياضي يمثل عددا كبيرا من العمليات والمعالجات الرياضية في تحليل الارتباطات بين المتغيرات (فقرات المقياس أو الاختبار) ومن ثم تفسير هذه الارتباطات واختزالها في عدد أقل من المتغيرات تدعى عوامل. ويساعد التحليل العاملي التوصل إلى أن السمة المقاسة أحادية البعد أو متعددة الأبعاد.

الصدق المتقاطع (عبر العينات) Cross Validation

عملية التحقق من النتائج المشتقة من تطبيق الاختبار على مجموعة من المفحوصين بإعادة تطبيقها على مجموعة أخرى مستقلة ولكن مشابهة للمجموعة الأولى.

الدرجة الزائية Z - Scores

هي درجة معيارية في توزيع متوسطه الحسابي يساوي (صفر) وإنحرافه المعياري يساوي (1). بذلك يمكن مقارنة الدرجة مع غيرها من الدرجات التي تنتمي لنفس التوزيع، وذلك بحساب الدرجة المعيارية الزائية لكل منها ومن ثم يكون التفضيل في ضوء مقدار قيمة الدرجة المعيارية (z).

الدرجة التائية T - Scores

هي درجة معيارية في توزيع متوسطه الحسابي يساوي (50) وانحرافه المعياري يساوي (10). وتمتد قيم الدرجات التائية للتوزيع الواحد بين (صفر) و (100).

الدرجة التساعية Stannine Scores

هي درجة معيارية في توزيع يتألف من تسع فئات، كل منها يسمى تساعي ويضم نسبة معينة من القيم (درجات أو صفات). ولا يخضع هذا التوزيع إلى قاعدة رياضية أو قانون إحصائي، بل يستند إلى معطيات تعتمد على مصلحة المؤسسة المستفيدة من هذه الدرجات.

الجماعة المرجعية (المعيارية) Reference Group

هي مجموعة الأفراد التي ينتسب الفرد إليها عند تفسير مستوى أدائه. والمفروض أن يتوفر في المجموعة المرجعية شروط الحداثة والمواءمة والتمثيل.

التحويل الخطي Linear Transformation

هو التحويل الذي يتم بإضافة ثابت إلى أو طرح ثابت من كل قيمة من قيم (درجات) التوزيع. أو ضرب كل قيمة أو قسمتها على ثابت، أو إجراء العمليات المذكورة مجتمعة. ولا يغير هذا النوع من التحويل من شكل توزيع البيانات الأصلي. ويدعى هذا التحويل بالخطي لأن العلاقة بين الدرجات الأصلية والدرجات المحولة تتخذ خطا مستقيما.

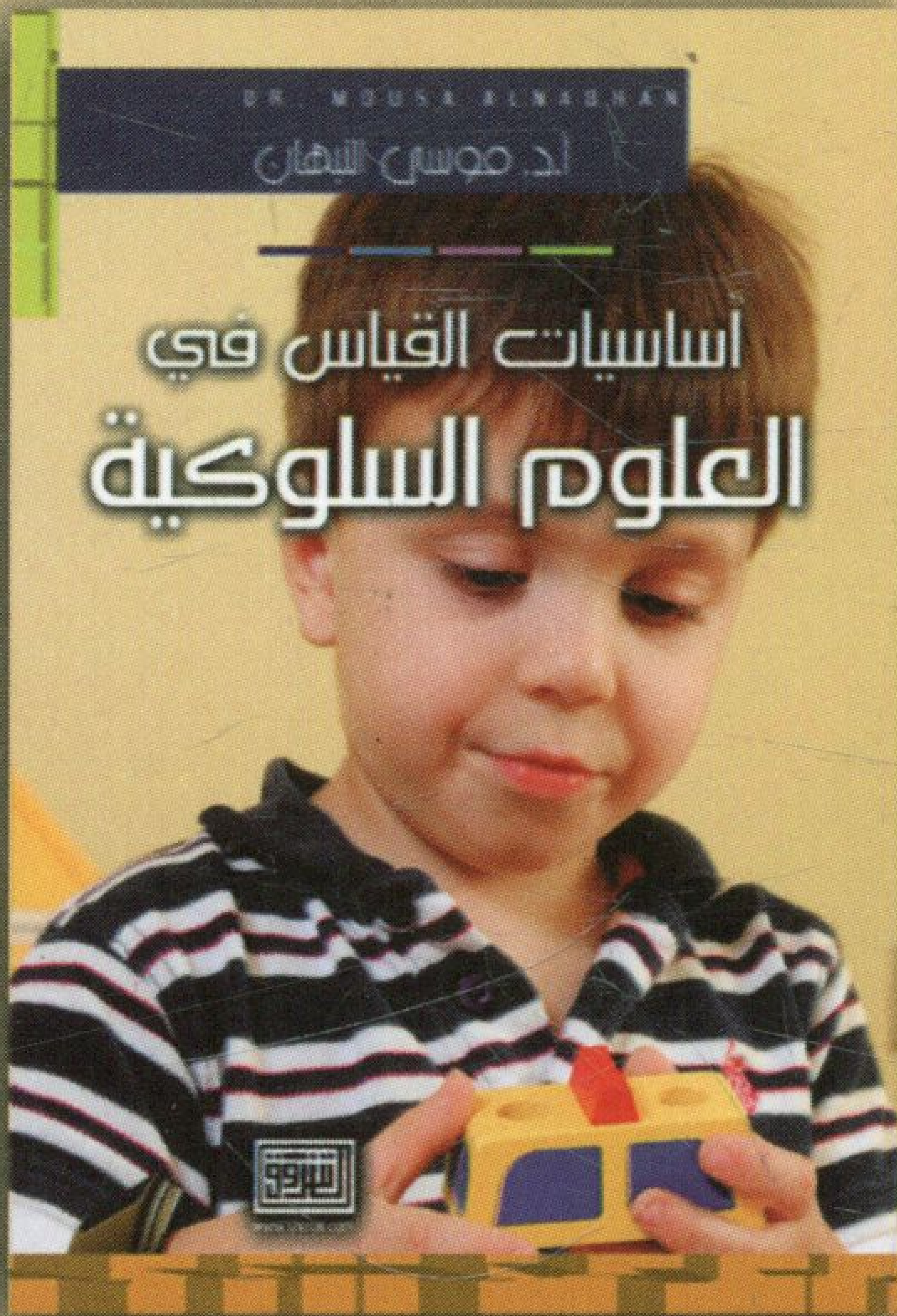
التحويل غير الخطي Non-Linear Transformation

هو التحويل الذي يتم عن طريق تحويل البيانات الأصلية إلى جذورها أو لوغاريتمها أو مقلوبها، ويعرف هذا النوع من التحويل بغير الخطي، لأن العلاقة بين البيانات الأصلية والبيانات المحولة ليست خطية. أي أن شكل توزيع البيانات بعد تحويلها تحويلا غير خطي يختلف عن شكلها قبل التحويل.

الاستكمال Extrapolation

هي عملية تفسير الدرجات التي تقع خارج نطاق المنحنى أو الجدول المعطى، سواء في حالة المعايير العمرية أو الدراسية أو غيرها.

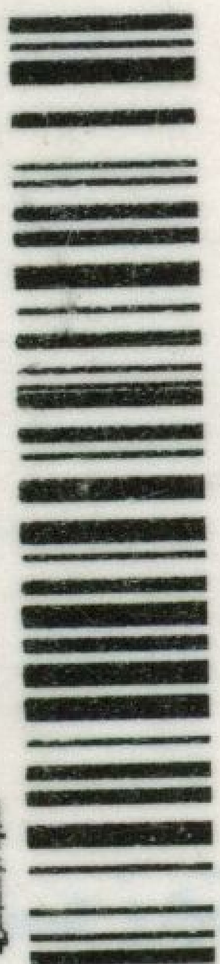
أساسيات القياس في
العلوم السلوكية



يحتل موضوع القياس أهمية مميزة في دراسة الظواهر والتنبؤ بتغيراتها. ويكون على أهمية أكبر عندما يتعلق بالعلوم السلوكية عامة، وبالسلوك الانساني تحديداً. ذلك بسبب طبيعة ذلك السلوك وتنوع وتشابك المتغيرات المؤثرة به.

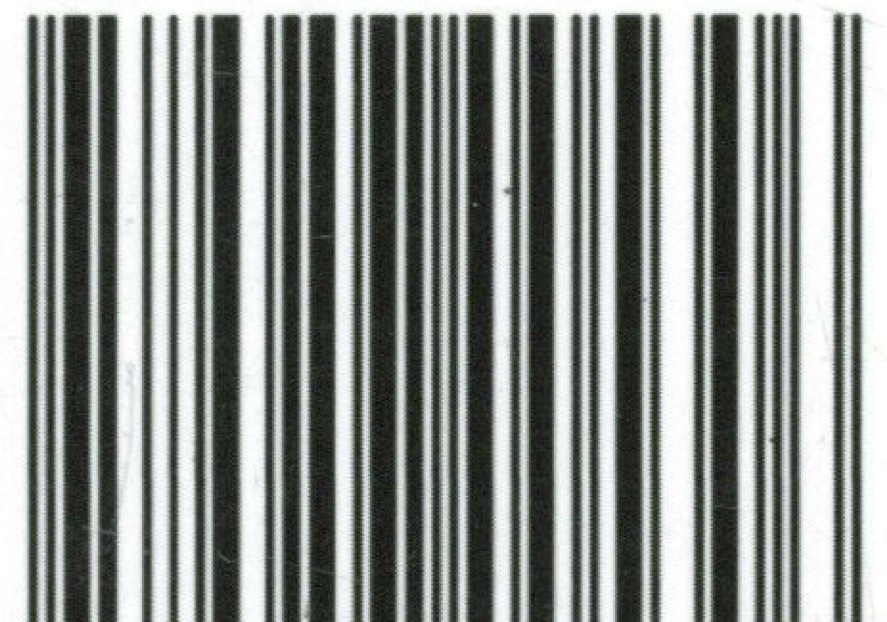
يجيء هذا العمل حلقة في سلسلة جهود، بدأت منذ زمن بعيد، عالجت القياس مفهوماً ومنهاجاً وأدوات. من هنا نعترف بأن هذا الجهد مكملاً لما سبق، مفيداً لما هو قائم، وليس اختراعاً أو اكتشافاً بأي حال من الأحوال.

ومهما يكن من أمر، فإن هذه الطبعة الثانية من هذا الكتاب تشكل إضافة نوعية، ربما تغطي فراغاً ما في المكتبة العربية. فهو حصيلة تدريس مواد تتعلق بالموضوع لأكثر من عشرين عاماً، وثمرة الإطلاع على عدد ليس بقليل من المصادر والمراجع الحديثة في مجالات الإحصاء والقياس والتقويم والتقييم.





دار الشروق للنشر والتوزيع
عمان-الأردن / رام الله - فلسطين


ISBN 9957-00-530-8
9789957005306